

Biologische Netzwerke

Metabolismus

3.1 Netzwerkrekonstruktion

Eine detaillierte Darstellung von Metabolismus mit den Mitteln der Graphentheorie ist die als gerichteter, gewichteter tripartiter Graph. Die Knoten repräsentieren in einer solchen Beschreibung Metaboliten, Reaktionen und Enzyme, die Kanten stehen für den Fluss von Stoffen, gewichtet mit den zugehörigen stöchiometrischen Koeffizienten, und katalytische Regulation (Albert, 2005). In dieser Repräsentation wird insbesondere die Regulation einer Reaktion durch mehrere Enzyme berücksichtigt. Eine Analyse der topologischen Eigenschaften von metabolischen Netzen legt die Vereinfachung dieser Strukturen zu unipartiten Graphen nahe: Metaboliten-Graphen (auch Substrat-Graphen genannt), Reaktions-Graphen und Enzym-Graphen lassen sich leichter handhaben, repräsentieren jedoch immer nur einen Teilaspekt der gesamten Netzwerkstruktur.

3.1.1 Substrat-Graphen

Wir beschäftigen uns in diesem Kapitel vor allem mit der Klasse der Substrat-Graphen, da sich diese intuitiv als strukturelle Basis für Transportprozesse interpretieren lassen (Wagner und Fell, 2001). Bei ihrer Konstruktion wird jedes Edukt einer Stoffwechselreaktion mit allen Produkten mit einem gerichteten Link verbunden, reversible Reaktionen werden mit zwei gegenläufigen Kanten beschrieben (siehe Abbildung 3.1). Im Mittelpunkt stehen dabei die von Ma und Zeng (2003b) aus der KEGG Datenbank (Kanehisa und Goto, 2000) erzeugten metabolischen Netze, die sich durch die sorgfältige Behandlung von Einzelfällen auszeichnen. Die Autoren erstellen eine Liste aller in der KEGG Datenbank enthaltenen Stoffwechselreaktionen und ergänzen die Information über Reversibilität und Richtung. Daraufhin verknüpfen sie jede Reaktion mit dem katalysierenden Enzym. Anzumerken ist, dass durchaus mehrere Reaktionen vom selben Enzym katalysiert werden können und eine Reaktion mehr als ein Enzym als Katalysator benötigen kann. Schließlich wird für jede Spezies die Verfügbarkeit des entsprechenden Enzyms überprüft: dazu wird die in der KEGG Datenbank enthaltene Genominformation nach der entsprechenden Sequenz durchsucht. Aus



Abb. 3.1. Die Repräsentation einer enzymatisch katalysierten Reaktion (links) als Substrat-Graph (rechts): Jedes Edukt wird mit jedem Produkt mit einer Kante verbunden.

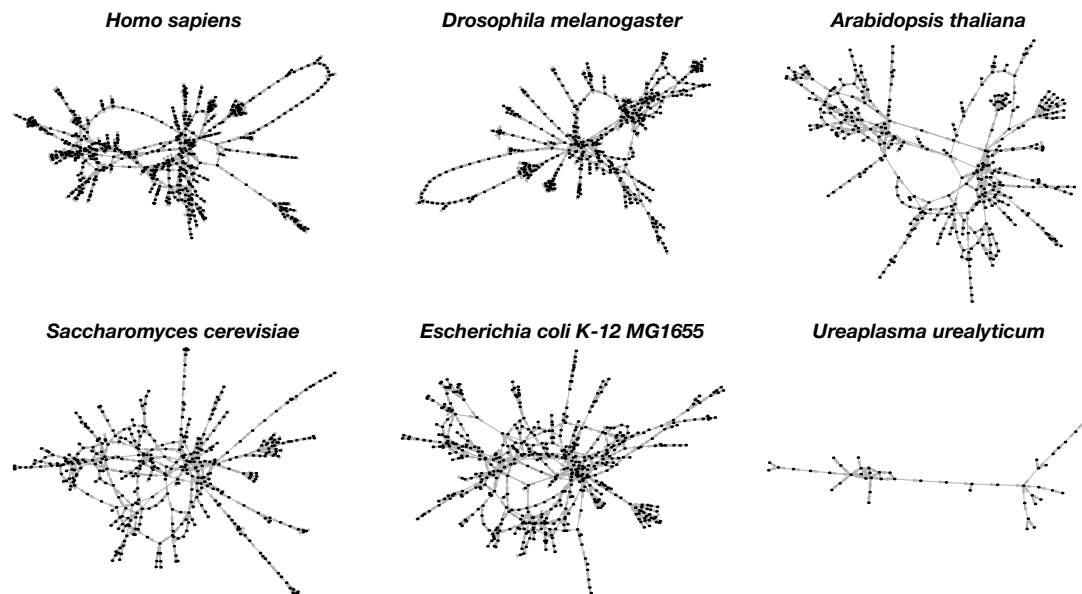


Abb. 3.2. Die größte zusammenhängende Komponente von sechs Substrat-Graphen aus der Sammlung von 107 metabolischen Netzen von Ma und Zeng (2003b): Kreise symbolisieren Metaboliten, Pfeile stehen für enzymatisch katalysierte biochemische Reaktionen. Der größte Graph mit 625 Knoten und 1061 Kanten ist das metabolische Netz des Menschen, *Homo sapiens*. Daneben sind die Graphen der Fruchtfliege *Drosophila melanogaster* und der Ackerschmalwand *Arabidopsis thaliana*, gezeigt. In der unteren Reihe sind die Substrat-Graphen der Bäcker-Hefe *Saccharomyces cerevisiae* und eines speziellen Stamms des Bakteriums *Escherichia coli* zu sehen. Der letzte Graph ist mit 49 Knoten und 82 Kanten das kleinste Netzwerk der Sammlung und gehört zu dem Bakterium *Ureaplasma urealyticum*.

den so gesammelten Reaktionsdaten werden die *current metabolites* (siehe auch Abschnitt 3.1.2), häufig auftretende Metaboliten, die als Träger von Elektronen oder funktionalen Gruppen dienen, entfernt und ein Netzwerk aus Substraten, verknüpft durch enzymatisch katalysierte Reaktionen, erstellt. Ma und Zeng rekonstruieren so die gerichteten Substrat-Graphen von 107 Spezies, von denen Abbildung 3.2 sechs ausgewählte Netze zeigt.

3.1.2 Strukturdebatte

Jeong et al. (2000) untersuchen die Topologie der Substrat-Graphen von 43 Spezies aus allen drei Domänen des Lebens – Eukaryoten, Prokaryoten und Archäen – und konstatieren zum einen eine skalenfreie Häufigkeitsverteilung der Grade eingehender und ausgehender Kanten, zum anderen eine extrem kurze mittlere Weglänge, etwa 3.2, für alle untersuchten Spezies, die zudem unabhängig von der Netzwerkgröße zu sein scheint. Diese Ergebnisse legen die Existenz von Hubs im Netzwerk nahe. Metaboliten, die an sehr vielen Stoffwechselreaktionen beteiligt sind, werden als *current metabolites* bezeichnet. Sie dienen in Reaktionen als Träger von Elektronen oder funktionalen Gruppen und umfassen, je nach Definition, ATP, ADP, NADH, NAD⁺, H₂O und Phosphat. Die Berücksichtigung dieser häufig auftretenden Substrate in metabolischen Netzen ist sinnvoll oder aber irreführend, je nach Blickwinkel und beabsichtigter Interpretation (Wagner und Fell, 2001). Werden die häufig auftretenden Substrate explizit aus den Stoffwechselreaktionen, in denen sie als *current metabolites* auftreten, entfernt, erhöht sich die mittlere Weglänge der zugehörigen Substrat-Graphen auf 9.57, 8.50 und 7.22 für Eukaryoten, Archäen und Bakterien (Ma und Zeng, 2003b).

Aufgrund der ersten topologischen Untersuchung zur Architektur metabolischer Netze (Jeong et al., 2000; Wagner und Fell, 2001; Ma und Zeng, 2003b), die – trotz der oben genannten Unterschiede in der Re-

konstruktion – allesamt die Small-world-Eigenschaft der zugrundeliegenden Graphen proklamierten, entspann sich eine Debatte über die angemessene Abstraktion von Metabolismus und die Interpretation der resultierenden Strukturen. Arita (2004) kritisierte frühere Arbeiten und die den Substrat-Graphen zugrundeliegenden Abstraktionsmethode und schlug stattdessen eine Netzwerkrekonstruktion vor, die sich auf die strukturelle Veränderung der Kohlenstoffkomplexe der Substrate konzentriert. Die dabei entstehenden Netze erhalten die tatsächlichen Weglängen der biochemischen Reaktionspfade, führen so auf große mittlere Weglängen und entbehren der Small-world-Eigenschaft. Anstelle von skalenfreien Graphen werden auch sogenannte *highly optimized tolerance* Modelle als adäquate Beschreibungen von Metabolismus diskutiert (Csete und Doyle, 2004; Tanaka, 2005). Bezüglich der angemessenen strukturellen Beschreibung der metabolischen Netze wird häufig das Bild der *bow-tie* (zu deutsch Fliege) verwendet (Ma und Zeng, 2003a; Csete und Doyle, 2004; Kitano, 2004): Stoffe aus der Umgebung, die Untermenge der Substrate, werden zerlegt und dem Kern-Metabolismus, dem Knoten der Fliege, zugeführt. Aus wenigen *Vorläufer-Metaboliten* werden dann all die Stoffe aufgebaut, die für den Erhalt der Zelle, ihr Wachstum und ihren Energiehaushalt unerlässlich sind. Diese werden dem Produkt-Untermenge zugeführt.

3.2 Dynamische Sonden

Die Dynamik einzelner Stoffwechselreaktionen hängt von den Konzentrationen aller beteiligten Komponenten ab. Ausgefeilte Modelle zur Beschreibung der Enzymkinetik wurden in den letzten Jahrzehnten entwickelt (siehe Klipp et al., 2005, für eine Übersicht). Eine vollständige Beschreibung aller metabolischen Abläufe ist wegen der Komplexität des Systems und aufgrund der unvollständigen Kenntnis der kinetischen Parameter unmöglich. Nichtsdestotrotz haben sich einige vollständige Beschreibungsmodelle etabliert, die eine zentrale Eigenschaft des metabolischen Netzes ausnutzen: die Konvergenz der Metabolitenkonzentrationen zu einem Fixpunkt, einem stabilen Zustand des dynamischen Systems. *Flux balance analysis* (kurz FBA) (Kauffman et al., 2003a; Palsson, 2006) beruht auf der stöchiometrischen Matrix, und damit auf der Topologie des metabolischen Netzwerks, auf einigen wenigen zusätzlichen Systemeigenschaften sowie der experimentell gemessenen Zusammensetzung der Biomasse eines Organismus. Diese und verwandte Fixpunkt-Methoden wurden in den vergangenen zehn Jahren erfolgreich zur Vorhersage phenotypischer Eigenschaften von Wildtyp-Spezies und Mutanten in verschiedenen Umgebungen eingesetzt (siehe beispielsweise Edwards und Palsson, 2000; Stelling et al., 2002; Famili et al., 2003).

Aufgrund des Erfolgs solcher Fixpunkt-Methoden wollen wir der Frage nachgehen, inwieweit die Architektur der metabolischen Netze dazu beiträgt, einen Fixpunkt der Stoffwechseldynamik in zuverlässiger Weise zu etablieren (Marr et al., 2007b). Um Fluktuationen einer komplexen Dynamik auf einem Netz zu dämpfen sind prinzipiell zwei Beiträge zu unterscheiden: lokale dynamische Effekte, die auf der speziellen Form der Enzymkinetik beruhen, und globale dynamische Effekt, die auf der Topologie der interagierenden Metaboliten beruhen.

Im Folgenden untersuchen wir den regularisierenden Einfluss der Netzwerkarchitektur, indem wir einfache, binäre Dynamiken gemäß dem in Abschnitt 2.1 eingeführten GCA-Formalismus auf Substrat-Graphen implementieren und das dynamische Verhalten der tatsächlichen Graphen mit dem randomisierter Pendants, d.h. verschiedener Nullmodelle vergleichen. Die Anwendung einer solchen dynamischen Sonde ergänzt eine topologische Debatte und verdeutlicht auf diese Weise ein zentrales Strukturmerkmal der metabolischen Netze: Das Dämpfen von Fluktuationen mit dem Ziel, einen stabilen, wiederherstellbaren und gleichzeitig auf Veränderungen der Umwelt sensiblen metabolischen Zustand des Systems zu etablieren. Daran schließt sich eine Überprüfung der Ergebnisse mit anderen Dynamiken, linearen Differentialgleichungssystemen und Phasenoszillatoren, an, die den regularisierenden Einfluss von Topologie auf komplexe Dynamiken bestätigt.

3.2.1 Zelluläre Automaten

Wir wenden die in Kapitel 2 entwickelten zellulären Automaten auf die metabolischen Netze von Ma und Zeng an. Die binären Dynamiken entsprechen dabei keineswegs der realen metabolischen Flussdynamik – wir wollen vielmehr untersuchen, ob im Mittel über alle Knoten Störungen im System eher gedämpft werden, uns so zu kleinen Entropiesignaturen führen, oder aber verstärkt werden, mit resultierenden binären Zeitreihen hoher Entropie. Einem Argument von Wagner und Fell (2001) folgend verwenden wir ungerichtete Varianten der Graphen, da sich Störungen prinzipiell in beide Richtungen einer enzymatisch katalysierten Reaktion ausbreiten können. Um Artefakte zu vermeiden und Vergleichbarkeit zwischen den Spezies zu gewährleisten betrachten wir ausschließlich die größte zusammenhängende Komponente der Netze. Diese umfasst bei Hefe 448 Knoten und 564 Kanten (siehe Abbildung 3.1). Wir wollen das Potential dieser Netze zur Regularisierung chaotischer und komplexer Dynamik untersuchen. Dazu wählen wir aus der Menge aller möglichen, einparametrischen GCAs solche Regeln aus, die zu maximalen Entropien, also komplexer Dynamik führen. Aufgrund der heterogenen Gradverteilung der Netze wählen wir einen vereinfachten GCA-Formalismus (siehe auch Abschnitt 2.4.1), indem wir in (2.3) $b = a$ setzen. Mit den beiden freien Parametern $(a, c) \in 0, 1, x, \bar{x}$ lassen sich, bei festem κ , $4^2 = 16$ verschiedene GCA-Regeln implementieren. Um alle einparametrischen GCA-Regeln einer vorgegebenen Architektur zu implementieren, wird κ zwischen 0 und 1 variiert. Die Intervallgröße, in denen eine Regel konstant bleibt, wird vom maximalen Grad des Graphen d^{\max} mit $1/d^{\max}$ vorgegeben. Aus der Menge aller möglichen einparametrischen GCA-Regeln, die sich durch die Variation von κ erzeugen lassen, wählen wir solche aus, die auf den metabolischen Netzen komplexe Dynamik hervorrufen. Für das Hefe-Netz sind dies lediglich zwei Regeln: (x, x, \bar{x}) mit $\kappa = 2/7$ sowie die symmetrische Entsprechung der Regel, (\bar{x}, \bar{x}, x) mit $\kappa = 5/7$. Abbildung 3.3 zeigt, dass sich diese beiden Regeln von einem topologischen Nullmodell, einem Hefe-Netz mit umgelegten Kanten unterscheiden. Im weiteren Verlauf dieses Kapitels untersuchen wir ausschließlich die GCA-Regel (x, x, \bar{x}) mit $\kappa = 2/7$, und bezeichnen diesen speziellen zellulären Automaten mit Ω :

$$\Omega : x_i(t+1) = \begin{cases} x, & \rho_i(t) \leq 2/7 \\ 1-x, & \rho_i(t) > 2/7 \end{cases} \quad (3.1)$$

In Abbildung 3.4a zeigt ein Raumzeitmuster die typische Entwicklung eines zufälligen Anfangszustands unter der Regel Ω . Die topologische Diversität dieses Graphen ist in der entstehenden Raumzeitentwicklung wiederzufinden: Konstante und streng oszillierende Zeitentwicklungen koexistieren mit chaotischen Zeitentwicklungen, wo Knoten zwischen alternierendem und konstantem Verhalten in unvorhersehbarer

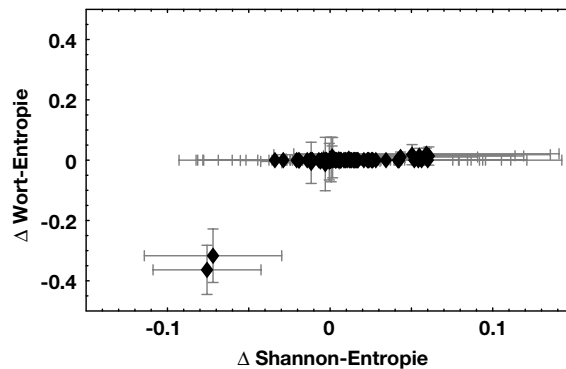


Abb. 3.3. Vergleich der Entropiesignaturen des Hefe-Substrat-Graphen mit seinem randomisierten Pendant. Nur zwei der $16 \cdot 15 = 240$ möglichen Regeln, durch Variation der Parameter $a = b$ und c sowie durch schrittweise Erhöhung von κ in Schritten von $1/d^{\max} = 1/14$ entstehen, führen zu einer deutlichen Entropiedifferenz. Die Fehlerbalken weisen auf die Streuung der Werte bei 10 verschiedenen Anfangsbedingungen und Graphenrealisierungen hin.

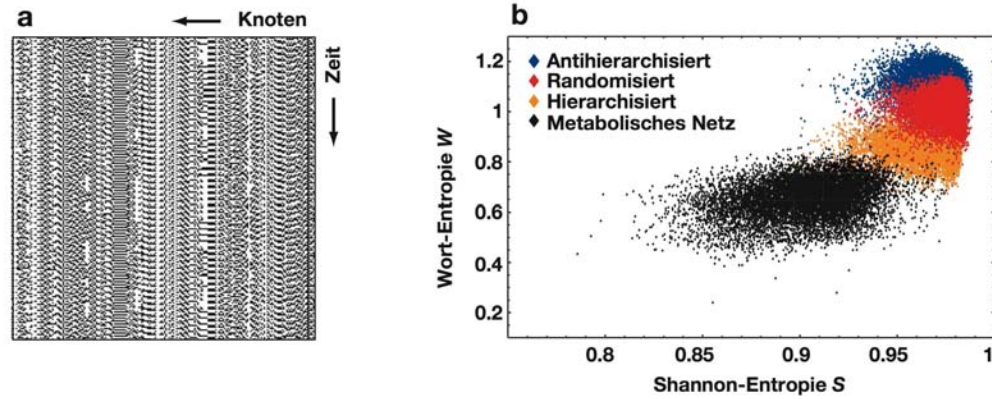


Abb. 3.4. (a) Das Raumzeitmuster der GCA-Regel Ω auf der größten zusammenhängenden Komponente des Substrat-Graphen von Hefe. Das Netz umfasst 448 Knoten und 564 Kanten, die Raumzeitmatrix zeigt 200 Zeitschritte von 200 Knoten. (b) Entropiesignatur des Hefe-Netzwerks (schwarz) sowie randomisierter (rot), hierarchisierter (orange) und antihierarchisierter Nullmodelle mit Grad-Grad-Korrelationen von 0.05, 0, +0.3 und -0.3. Jeder Punkt in der Entropieebene repräsentiert die Entropiesignatur des entsprechenden Netzwerks für eine zufällig ausgewählte Anfangsbedingung. 1000 Anfangsbedingungen pro Netzwerktyp sind gezeigt.

Weise wechseln. Den Vergleich der Auswertung solcher Muster mit drei Nullmodellen, in denen das metabolische Netzwerk randomisiert, hierarchisiert und antihierarchisiert (siehe Kapitel 1.4) wurde, zeigt Abb. 3.4b: Die Entropien des realen Graphen sind im Vergleich zu den Nullmodellen im Mittel stark reduziert, was auf eine Regularisierung der durch Ω implementierten dynamischen Sonde hinweist. Die reale Systemarchitektur zeichnet sich also gegenüber sehr ähnlichen (die Gradsequenz der Graphen bleibt erhalten), leicht modifizierten Architekturen in ihrer Antwort auf die auferlegte Dynamik deutlich aus.

Wir wollen überprüfen, ob diese Netzwerkeigenschaft bei allen metabolischen Netzen von Ma und Zeng zu beobachten ist. Dazu implementieren wir Ω auf allen Substrat-Graphen und den entsprechenden randomisierten Nullmodellen. Für 99 der 107 Spezies in der Datenbank finden sich reduzierte Entropiesignaturen.

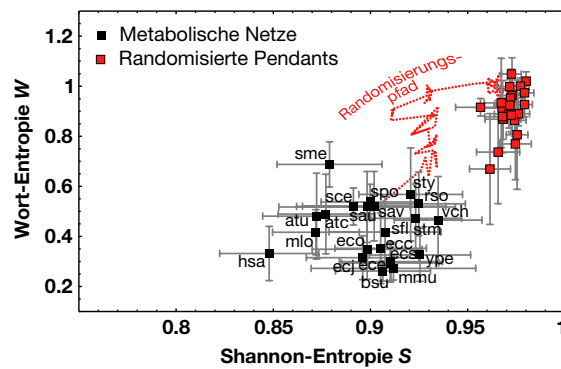


Abb. 3.5. Mittlere Entropiesignatur der 22 metabolischen Netze mit mittlerem Grad $L/N > 1.22$ und ihren randomisierten Nullmodellen. Die Region der metabolischen Netze (schwarz) ist deutlich getrennt von den randomisierten Pendants (rot) gleicher Größe, identischem Vernetzungsgrad und gleicher Gradsequenz. L Randomisierungsschritte wurden dabei auf jedes Netz angewandt. Fehlerbalken verweisen auf die Abweichung der Entropie Signaturen aufgrund verschiedener zufälliger Anfangsbedingungen, bzw. verschiedener Nullmodell Realisierungen. Schrittweise Randomisierung des Hefe Netzes führt auf einen Pfad (gepunktete Linie), hin zu höheren Entropien. Die Abkürzungen sind den Spezies in Tabelle 3.1 zugewiesen.

Tabelle 3.1. Einige Observablen der Netzwerke der 22 Spezies aus Abbildung 3.5. Die Tabelle zeigt die Zahl der Knoten N , den mittleren Grad L/N , die mittlere Weglänge PL , den Cluster-Koeffizienten CC sowie die Differenz der Entropiesignaturen des metabolischen Netzes und seines randomisierten Nullmodells.

Spezies	Abkürzung	Domäne	N	L/N	CC	PL	ΔS	ΔW
<i>H. sapiens</i>	hsa	Eukaryot	625	1.25	-0.09	11.3	0.13	0.60
<i>M. musculus</i>	mmu	Eukaryot	320	1.23	-0.10	10.7	0.05	0.40
<i>S. cerevisiae</i>	sce	Eukaryot	448	1.26	0.05	9.47	0.09	0.50
<i>S. pombe</i>	spo	Eukaryot	337	1.25	0.07	9.36	0.08	0.46
<i>A. tumefaciens</i> C58 (Cereon)	atc	Prokaryot	504	1.22	0.07	9.01	0.09	0.44
<i>A. tumefaciens</i> C58 (UWash/Dupont)	atu	Prokaryot	502	1.22	0.07	9.02	0.10	0.40
<i>B. subtilis</i>	bsu	Prokaryot	501	1.22	0.04	9.17	0.07	0.54
<i>E. coli</i> K-12 MG1655	eco	Prokaryot	563	1.26	0.04	8.76	0.07	0.39
<i>E. coli</i> K-12 W3110	ecj	Prokaryot	562	1.26	0.04	8.83	0.07	0.68
<i>E. coli</i> O157 EDL933	ece	Prokaryot	545	1.27	0.04	8.68	0.07	0.68
<i>E. coli</i> O157 Sakai	ecs	Prokaryot	548	1.27	0.04	8.71	0.06	0.66
<i>E. coli</i> CFT073	ecc	Prokaryot	484	1.25	0.01	8.94	0.07	0.60
<i>M. luti</i>	mlo	Prokaryot	553	1.23	0.08	10.2	0.10	0.51
<i>R. solanacearum</i>	rso	Prokaryot	505	1.23	0.05	10.3	0.05	0.36
<i>S. aureus</i> N315 (MRSA)	sau	Prokaryot	398	1.22	0.06	8.90	0.07	0.37
<i>S. aureus</i> Mu50 (VRSA)	sav	Prokaryot	397	1.22	0.06	8.90	0.05	0.39
<i>S. flexneri</i>	sfl	Prokaryot	518	1.25	0.03	8.83	0.06	0.47
<i>S. meliloti</i>	sme	Prokaryot	563	1.24	0.06	10.0	0.09	0.36
<i>S. typhimurium</i>	stm	Prokaryot	582	1.24	0.05	10.3	0.05	0.39
<i>S. typhi</i>	sty	Prokaryot	515	1.27	0.04	8.53	0.05	0.34
<i>V. cholerae</i>	vch	Prokaryot	419	1.26	0.07	8.60	0.04	0.30
<i>Y. pestis</i> CO92	ype	Prokaryot	514	1.23	0.02	9.08	0.04	0.54

Die Netzwerke der verbleibenden 8 Spezies sind extrem schwach vernetzt und daher kaum zu randomisieren. Wir wählen aus den 99 Spezies die Netze mit einem mittleren Grad $L/N > 1.22$ aus. Die gemittelten Entropiesignaturen der 22 Graphen, die diese Bedingung erfüllen, lassen sich statistisch robust auswerten. Sie bilden einen Cluster in der Entropieebene. Abbildung 3.5 zeigt die Entropiesignaturen dieser Netze zusammen mit ihren randomisierten Nullmodellen und einem Pfad, der das systematische Anwachsen der Entropien mit der Randomisierungstiefe für das metabolische Netz von Hefe verdeutlicht. Der umgekehrte Weg, in Richtung erniedrigter Entropiesignaturen, kann im Prinzip durch einen entsprechenden Randomisierungsalgorithmus realisiert werden. In diesem Sinne könnte die Entropieebene als Orientierung für die simulierte Evolution von Netzen mit maximaler Regularisierungskapazität dienen. Tabelle 3.1 listet die Speziesnamen zusammen mit einigen topologischen Observablen und den Differenzen der Entropiesignaturen auf.

3.2.2 Minimale Modifikationen

Das Randomisierungsprotokoll zur Erzeugung der entsprechenden Nullmodelle verändert eine Reihe zentraler topologischer Eigenschaften: Die mittlere Weglänge wird verkürzt, der Cluster-Koeffizient und Grad-Grad-Korrelationen werden reduziert, modulare Substrukturen des Graphen werden durch das Umlegen von Kanten zerstört. Um zu überprüfen, ob die Optimierung der metabolischen Netze allein mit der Veränderung dieser Observablen zu erklären ist oder aber über das Konzept topologischer Beschreibungsgrößen hinausgeht, analysieren wir den Effekt von minimalen topologischen Störungen. Durch minimale Eingriffe lässt sich eine strukturelle Optimierung der realen Netze unter Beibehaltung wichtiger topologischer Kenngrößen untersuchen. Dazu bestimmen wir die Entropiesignatur des Hefe-Netzwerks und eines modifizierten Netzes, das sich vom Originalnetz nur durch einen einzigen Randomisierungsschritt unterscheidet,

für verschiedene zufällige Anfangsbedingungen (Abbildung 3.6a). Der Mittelwert der Entropieänderung ist spezifisch für den durchgeführten Modifikationsschritt (Abbildung 3.6b). Eine Mittelung über viele verschiedene Modifikationen führt letztlich auf die mittlere Änderung der Entropiesignatur (Abbildung 3.6c), charakteristisch für das angewandte Modifikationsprotokoll. Wir wenden minimale topologische Modifikationen auf das Hefe-Netzwerk an und analysieren die Änderung der Entropiesignatur mit den eben beschriebenen Mitteln (Abbildung 3.6d) für Randomisierung, Hierarchisierung, Antihierarchisierung, Randomisierung unter Erhaltung des Cluster-Koeffizienten sowie der mittleren Weglänge und Randomisierung ausschließlich inter- sowie intramodularer Kanten. Die entsprechenden Module im Hefe-Netz wurden mit einem Weglängen-Algorithmus, wie in Ma et al. (2004b) beschriebenen, bestimmt und sind in Abbildung 3.7 dargestellt. Ein mittleres Anwachsen der Entropiesignatur ist für jedes Modifikationsprotokoll offensichtlich (Abbildung 3.6d). Dabei wurde stets der Zusammenhang der modifizierten Graphen, sowie ein ausreichend großer Möglichkeitsraum für das Sampling von Nullmodellen gewährleistet. Das Anwachsen der Änderung der Entropiesignatur mit den Modifikationsschritten zeigt Abbildung 3.8 für den einfachen Randomisierungsprozess. Neben den Mittelwerten ist hier auch die Schar von Vektoren, die aus den 200 verschiedenen Einzelmodifikationen resultiert, angegeben. Der Extremfall sehr stark hierarchisierter und antihierarchisierter Graphen wurde schon in Abbildung 3.4b gezeigt.

Das Anwachsen der Entropieänderung unter jeder topologischen Störung zeigt, dass die bemerkenswerte Eigenschaft metabolischer Netze, die deutliche Regularisierung chaotischer Dynamik, nicht mit einer trivialen topologischen Observablen zu erklären ist. Darüberhinaus besteht diese Eigenschaft jenseits einfacher Grad-Grad-Korrelationen und modularer Substrukturen. Ein möglicher Einwand gegen die beobachteten Besonderheiten der metabolischen Netze ist die Repräsentation als Substrat-Graph. Wir haben die regularisierende Kapazität metabolischer Netze daher für andere Netzwerkrepräsentationen untersucht: 43 unipartite Substrat-Graphen, in denen die current metabolites nicht entfernt wurden (Jeong et al., 2000), gerichtete Repräsentanten der Ma/Zeng Graphen, sowie ein Enzym-Graph, der aus der stöchiometrischen

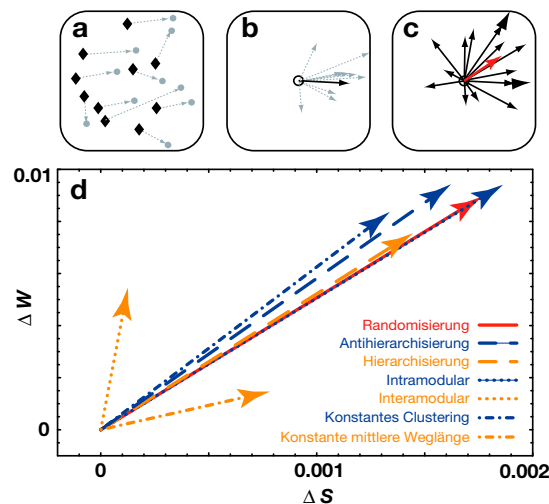


Abb. 3.6. Der Effekt minimaler topologischer Störungen im Hefe-Netzwerk. Die Methode: (a) Wir berechnen die Entropiesignatur für den Originalgraphen (Rauten) und ein minimal modifiziertes Pendant (Kreise) für verschieden Anfangsbedingungen. (b) Wir mitteln (schwarzer Pfeil) über die Anfangsbedingung (graue Pfeile). (c) Wir mitteln (roter Pfeil) über verschiedene minimale Modifikationen (schwarze Pfeile). Ergebnis (d): Die Entropiesignatur wird durch jede der angewendeten minimalen Modifikationen systematisch erhöht. Die verschiedenen Protokolle randomisieren (rot), hierarchisieren (gelb, gestrichelt) und antihierarchisieren (blau, gestrichelt) das Hefe-Netz. Gepunktete Pfeile zeigen die ausschließliche Randomisierung von Kanten innerhalb (blau) und zwischen (gelb) den Modulen des Netzes. Die verbleibenden Pfeile zeigen Randomisierungsprotokolle unter Erhaltung des Cluster-Koeffizienten (blau) und der mittleren Weglänge (gelb) im Graphen.

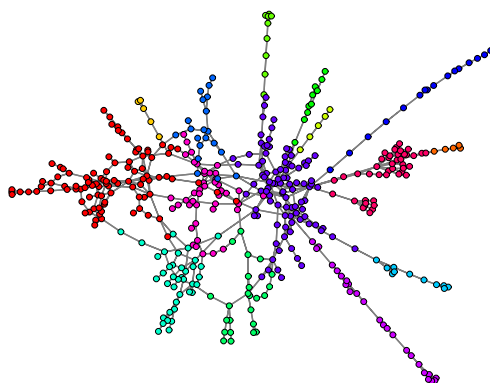


Abb. 3.7. Die Module des Hefe-Substrat-Graphen, bestimmt mit dem Algorithmus aus Ma et al. (2004b), dem die Weglänge zwischen zwei Knoten zugrundeliegt.

Matrix von Hefe¹ erstellt wurde, zeigen allesamt erniedrigte Entropiesignaturen im Vergleich zu randomisierten Nullmodellen. Aus diesem Grund nehmen wir an, dass die beobachtete regularisierende Kapazität eine generische Eigenschaft der Architektur des Metabolismus ist, jenseits der spezifischen Repräsentation.

3.2.3 Modellgraphen

Der Vergleich der Substrat-Graphen mit randomisierten Nullmodellen zeigt eine Auszeichnung der realen Graphen gegenüber verschiedenen Untermengen von Graphen mit der gleichen Gradsequenz. Ein Vergleich mit den in Abschnitt 1.3 eingeführten Modellgraphen führt über diese Menge hinaus. Wir setzen daher Small-world-Graphen und skalenfreien Graphen mit der gleichen Zahl von Knoten und Kanten wie die entsprechenden 22 Substrat Graphen der dynamischen Sonde Ω aus und analysieren die dynamische Antwort dieser synthetischen Architekturen. Abbildung 3.9 zeigt die Entropiesignatur der 22 Substrat-Graphen zusammen mit den jeweiligen Modellgraphen und randomisierten Pendants. Bemerkenswert ist zunächst der Grad an Spezifität, den unsere dynamische Sonde erzielt: Die verschiedenen Graphentypen clustern untereinander und separieren voneinander deutlich in der Ebene, Small-world-Graphen mit deutlich erhöhten Entropiewerten, skalenfreie Graphen mit einer leicht erniedrigten Wort-Entropie. Offensichtlich reagieren die beiden Entropien sehr unterschiedlich auf die strukturellen Unterschiede der Netzwerktypen. Die Entropieänderung aufgrund der Randomisierung misst die Abweichung der Modellgraphen von den entsprechenden Nullmodellen mit gleicher Gradsequenz. Für skalenfreie Graphen ist diese Abweichung klein:

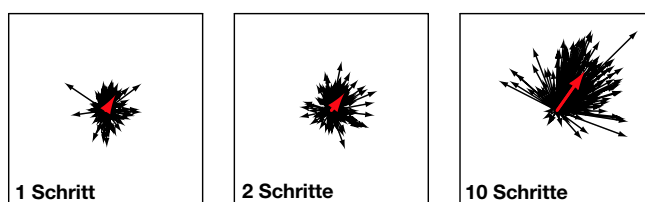


Abb. 3.8. Änderung der Entropiesignatur des Hefe-Netzes durch Randomisierungsprotokolle unterschiedlicher Tiefe. Stets wird die Entropiesignatur im Mittel erhöht (rote Pfeile). Bei zehn Randomisierungsschritten (rechts) ist eine Vorzugsrichtung der Signaturänderung der 200 modifizierten Graphen klar zu erkennen (schwarze Pfeile).

¹ Die stöchiometrischen Matrizen von Hefe und weiteren Organismen, erstellt von der Palsson-Gruppe sind frei zugänglich unter <http://gcrp.ucsd.edu/organisms/index.html>

Lediglich die Wort-Entropie der randomisierten Nullmodelle ist leicht reduziert. Für Small-world-Graphen dagegen bedeutet Randomisierung eine Auflösung der lokalen Nachbarschaftsstruktur, eine dramatische Reduktion der Cliquenhaftigkeit des Netzwerks. Dementsprechend stark weichen die randomisierten Graphen hier von den synthetischen Ausgangsgraphen ab. Die Reduktion der Entropiesignatur durch Randomisierung der Modellgraphen beweist darüberhinaus, dass die minimale Signatur der realen metabolischen Netze kein trivialer Effekt des Randomisierungsprotokolls ist.

3.2.4 Einfache lineare Differentialgleichungen

Einfache lineare Differentialgleichungen (DGLs) approximieren aufgrund des kontinuierlichen Zustandsraumes in etwas realistischerer Weise die Dynamik metabolischer Reaktionen. Sie lassen sich als Entwicklungen erster Ordnung der realen Dynamik um den Fixpunkt herum interpretieren. Wir betrachten also ein System von gekoppelten linearen DGLs, wobei die Struktur der Kopplung durch die Topologie des metabolischen Netzes und die Stärke der Kopplung durch zufällig ausgewürfelte Parameter festgelegt ist. Wir wollen im Rahmen linearer DGLs untersuchen, wie sich Störungen im System auf das Erreichen des stabilen Zustandes auswirken, inwieweit diese Systemeigenschaft von der Topologie geprägt wird und ob die mit der binären dynamischen Sonde gemessenen Entropiesignaturen mit den Propagationseigenschaften solcher Störungen korrelieren.

Die Zeitentwicklung der Knoten im System folgt einer einfachen linearen DGL 1. Ordnung,

$$\dot{\vec{x}}(t) = \vec{x}_0 + C \vec{x}(t), \quad (3.2)$$

mit einer Kopplungsmatrix C . wir können ohne Beschränkung der Allgemeinheit $\vec{x}_0 = 0$ setzen. Damit ist dem System der Fixpunkt $\vec{x}(t) = 0$ eingeschrieben. Das System befindet sich in einem Fixpunkt, wenn $\vec{x} = 0 \Leftrightarrow C \vec{x} = 0$. Dieser Fixpunkt ist dann stabil, wenn die Realteile aller Eigenwerte der Jacobi-Matrix des Systems negativ sind. Im vorliegenden linearen System ist die Jacobi-Matrix gerade die Kopplungsmatrix C . Wir bauen C folgendermaßen auf: Jeder Kopplungsterm der Adjazenzmatrix erhält einen zufälligen Faktor w_{ij} , um die Stärke der Kopplung von Simulation zu Simulation zu variieren, $\tilde{A}_{ij} = w_{ij} A_{ij}$. Darüberhinaus stellen wir jeden Knoten mit dem gleichen negativen Dämpfungsterm s in der Kopplungsmatrix aus, um die Existenz eines stabilen Fixpunktes zu vereinfachen. Die Kopplungsmatrix ist dann von der Form:

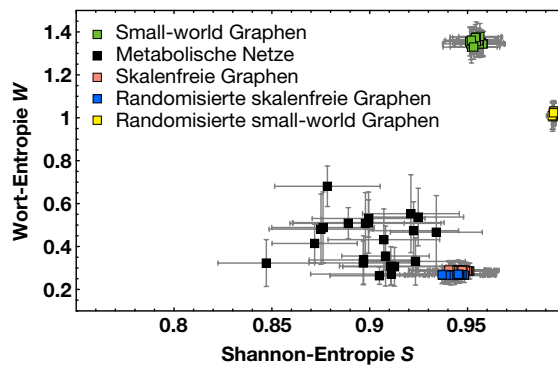


Abb. 3.9. Mittlere Entropiesignatur der 22 Spezies aus Abbildung 3.5, zusammen mit synthetischen Graphen gleicher Größe und Vernetzung. Wir zeigen die Signaturen für die realen Netze (schwarz), skalenfreie Graphen (rosa), randomisierte skalenfreie Graphen (blau), Small-world-Graphen (grün) und randomisierte Small-world-Graphen (gelb). Die Small-world-Graphen wurden durch das Umlegen von 5% der Kanten einer regulären Kette generiert. Die randomisierten Netzen wurden durch das Tauschen von L Paaren von Knoten der Ausgangsgraphen erzeugt. Wir erzeugen 100 synthetische Netze und mitteln jeweils über 100 Anfangsbedingungen.

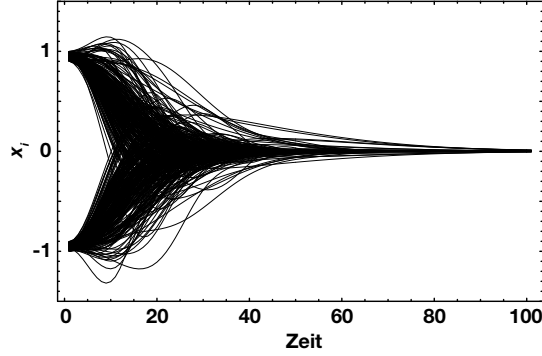


Abb. 3.10. Transient des linearen Differentialgleichungssystems aus Gl. (3.2) auf dem gerichteten Substrat-Graphen des metabolischen Netzes von Hefe mit zufällig gewählten positiven und negativen Anfangsbedingungen $\pm x_i(0) \in [0.9, 1]$. Die einzelnen Zeitentwicklungen x_i erreichen verschieden schnell den stabilen Fixpunkt des Systems.

$$C = s \mathbf{1} + \tilde{A}, \quad (3.3)$$

wobei $\mathbf{1}$ die Identitätsmatrix bezeichnet.

Der Einfachheit halber werden keine Störungen per Hand eingefügt, vielmehr untersuchen wir den *Transienten* des Systems, also die Zeit, die die dynamischen Einzelemente benötigen, um von zufällig gewählten Anfangsbedingungen $\vec{x}(0)$ aus den stabilen Fixpunkt des Systems zu erreichen. Wir müssen eine Reihe von Parametern wählen, um das System ausreichend zu spezifizieren. Für die im Folgenden dargestellten Ergebnisse wählen wir $s \in [-1.1, 1]$, $w_{ij} \in [-1, 1]$ und $x_i(0) \in [-1, -0.9] \cup [0.9, 1]$, wobei die exakte Größe und Position der Intervalle keine fundamentale Rolle spielt. Abbildung 3.10 zeigt den Transienten einer so implementierten Dynamik mit zufälligen Parametern auf dem gerichteten metabolischen Netzwerk von Hefe, genauer dem Substrat-Graphen aus der Datenbank von Ma und Zeng (2003b). Jede Trajektorie entspricht der Zeitentwicklung $x_i(t)$ eines Knotens im Netzwerk.

Wir berechnen für jeden Knoten i die Transientenzeit, d.h. die Zeit, für die x_i größer als ein gewählter Schwellenwert ist. Die Transientenzeit des Ensembles von dynamischen Einheiten lässt sich mit einer statistischen Größe der so berechneten Einzeltransienten angeben. Die Wiederholung eines solchen Experiments für viele verschiedene Parameterrealisierungen bei gleicher Topologie, sowie für verschiedene Realisierungen der Topologie lässt auf den Einfluss der Architektur auf die Transientenzeit des Netzwerks schließen. In Abbildung 3.11a ist der über zehn Durchläufe und hundert verschiedene Graphenrealisationen gemittelte Median der Transientenzeit des Ensembles gegen die Randomisierungstiefe gezeigt, wobei wie zuvor die

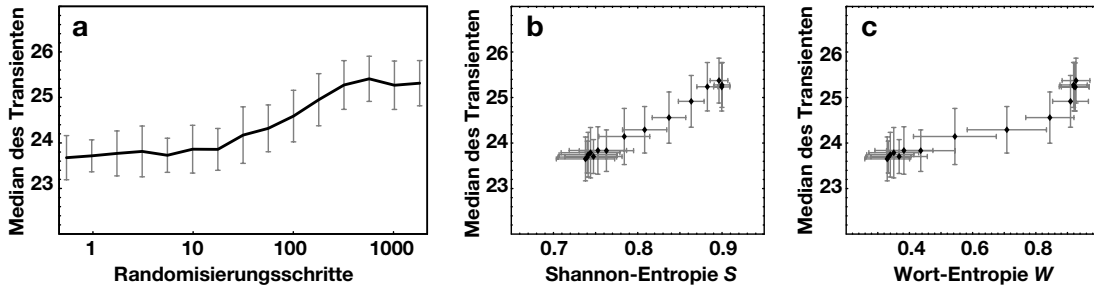


Abb. 3.11. (a) Median des Ensembles von Transientenzeiten, gemittelt über 100 Graphenrealisationen und je 10 Durchläufen mit zufälligen Parametern. Der mittlere Median steigt deutlich mit der Anzahl der Randomisierungsschritte. (b) und (c) Korrelation des Medians mit der Entropiesignatur der zellulären Automaten Dynamik $\Omega : (x, x, \bar{x})$, $\kappa = 2/7$ aus Abschnitt 3.2.1.

Gradsequenz des Graphen bei solchen Operationen erhalten bleibt. Offensichtlich steigt die Transientenzeit mit dem Randomisierungsgrad des Hefe-Netzwerks an, was auf die erhöhte Kapazität der realen Topologie zur Dämpfung komplexer Dynamik hindeutet. Abbildungen 3.11b und 3.11c zeigen die Korrelationen der gemittelten Mediane der Transientenzeit mit den mittleren Entropiewerten der Entropiesignatur. Die dynamischen Sonde Ω bringt also die gleichen Signaturen hervor wie ein System linearer DGLs, dass an die tatsächliche Dynamik deutlich stärker angepasst ist. Zwei Vorteile der nichtlinearen Sonde lassen sich sofort angeben: Zum einen entbehrt diese Darstellung eine Verortung der Simulation in einem hochdimensionalen Parameterraum, wie es für die linearen DGLs nötig ist. Zum anderen lassen sich mit der Methode Netze verschiedenster Herkunft parallel untersuchen, ohne dass durch die Auswahl einer bestimmten Dynamik bereits eine spezielle Architektur ausgezeichnet ist.

Inwieweit hängen die individuellen Transientenzeiten von den topologischen Eigenschaften der Knoten ab? Wir untersuchen dazu die Korrelation zwischen mittlerer Transientenzeit T_i und der Zahl der eingehenden und ausgehenden Knoten in Abbildung 3.12. Während die Zahl der ausgehenden Kanten erwartungsgemäß keinen Einfluss auf die Transientenzeit des entsprechenden Knoten hat, zeigt sich ein leichter Anstieg von T_i mit der Zahl der eingehenden Knoten. Die Kopplung zu benachbarten Knoten verhindert also ein schnelles Auffinden des individuellen Fixpunktes der Knoten. Dieser topologische Effekt spielt beim Vergleich der realen Struktur mit randomisierten Varianten keine Rolle, da die Gradsequenz in allen Graphen ja die gleiche ist. Der kooperative Effekt der metabolischen Netzwerkarchitektur, der zu kurzen Transientenzeiten und damit einem schnellen Erreichen des stabilen Fixpunktes des Systems führt, bleibt von diesem Ergebnis also unbeeinflusst. Dieser Effekt ist erst dann nachzuweisen, wenn die individuellen Observablen – wie die Transientenzeit, oder die Entropiesignatur – zusammen betrachtet werden. Auf dieser systemischen Ebene offenbart sich der Einfluss der Topologie auf eine implementierte Dynamik.

3.2.5 Erweiterungen

Der Vergleich verschiedener Topologien durch gemittelte Entropiesignaturen zeigt die erhöhte Regularisierung komplexer Dynamiken auf realen metabolischen Netzen. Dabei bleiben die lokalen Entropiedifferenzen der einzelnen Knoten unberücksichtigt. Motiviert durch die Ergebnisse aus den vorhergehenden Abschnitten ist eine Erweiterung der Untersuchung metabolischer Netze mit zellulären Automaten als dynamischen Sonden in mehrere Richtungen denkbar. Zum einen lässt sich anhand der über viele Anfangsbedingungen gemittelten, individuellen Knotenentropien eine Entropie-Topographie erstellen, die den Umgang der einzelnen Knoten mit Störungen beschreibt. Knoten mit kleinen Entropien weisen auf Metaboliten hin, deren Konzentration robust gegenüber Schwankungen ihrer Nachbarmetaboliten ist. Diese Gruppe konstituiert wahrscheinlich essentielle Transportwege sowie rauschunempfindliche Reaktionen aus dem Kern der

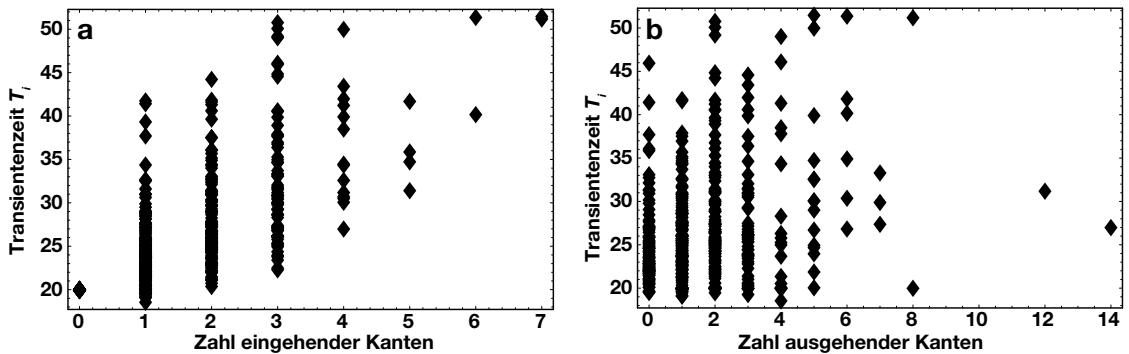


Abb. 3.12. Individuelle Transientenzeit T_i des Knotens i , gemittelt über 50 Realisationen mit zufälligen Parametern, gegen die Zahl der eingehenden (a) und ausgehenden (b) Kanten im gerichteten metabolischen Netz von Hefe.

Netze (Almaas et al., 2005). Knoten mit hohen Entropien dagegen greifen Störungen aus ihrer Nachbarschaft auf und prozessieren diese weiter. Diese Knoten könnten im Netz als Sensoren dienen, die zwischen verschiedenen funktionalen Zuständen des Netzes, beispielsweise bei veränderten Umweltbedingungen, oder im Verlauf des Zellzyklus, schalten, indem sie auf Veränderungen reagieren und diese weitergeben.

Interessant wäre auch die Untersuchung der Verbindung zwischen Entropie und der Reorganisation von Flüssen bei genetischen Mutationen und der Anpassung an neue Umweltbedingungen. Die FBA sucht im Flussraum stets nach der Lösung, die die Zielfunktion maximiert. Neuere Studien zeigen, dass sich der tatsächliche Flussvektor bei Mutationen an der Lösung des Wildtyps orientiert. Teilaspekte der Restrukturierung der Flüsse werden von der Methode der minimalen Modifikation der metabolischen Flüsse, MO-MA (Segré et al., 2002), andere mit einer Minimierung der signifikanten Flussänderungen, ROOM (Shlomi et al., 2005) besser erfasst. Die Reorganisation von Flüssen ist also ein nichttrivialer dynamischer Prozess, der über die Möglichkeiten der FBA hinausgeht und dem die Topologie des Netzes klar aufgeprägt ist. Es bleibt zu untersuchen, inwiefern die individuellen Entropien der Knoten beim Auffinden solcher Reorganisationspfade behilflich sein können.

Genregulation

4.1 Das Transkriptions-Regulations-Netzwerk von *E. coli*

Das Bakterium *Escherichia coli* ist einer der wichtigsten prokaryotische Modellorganismen der molekularbiologischen/-genetischen Forschung. *E. coli* Kulturen wachsen schnell und robust, bei optimalen Bedingungen verdoppelt sich die Bakterienzahl innerhalb von 20 Minuten. Die Bakterien sind Stäbchenförmig, etwa 1 μm breit und bis zu 6 μm lang. Seit Jahrzehnten wird der Metabolismus von *E. coli* studiert, gentechnisch veränderte Stämme werden industriell genutzt, beispielsweise zur Insulinproduktion. Das Genom von *E. coli* wurde bis 1997 vollständig sequenziert, es ist 4.6 Megabasenpaare lang und beinhaltet etwa 4500 Gene. Viele frei zugängliche Datenbanken beschäftigen sich mit verschiedenen Aspekten des Bakteriums. Die Regulon-Datenbank¹ (RegulonDB) (Salgado et al., 2006) enthält in ihrer aktuellen Fassung (5.6) 2822 regulatorische Wechselwirkungen zwischen 1345 *E. coli* Genen. Vermittelt wird diese Art der Regulation durch Genprodukte, die aus der Kopie der DNA (Expression), und der Überführung dieser Kopie in ein Protein entstehen. Proteine mit einer regulatorischen Wirkung heißen Transkriptionsfaktoren (TFs) und können an die Promoterregionen anderer Gene binden. Dadurch wird die Expression dieser Gene beeinflusst. Neben den Namen der Gene und Genprodukte können in RegulonDB Informationen über die

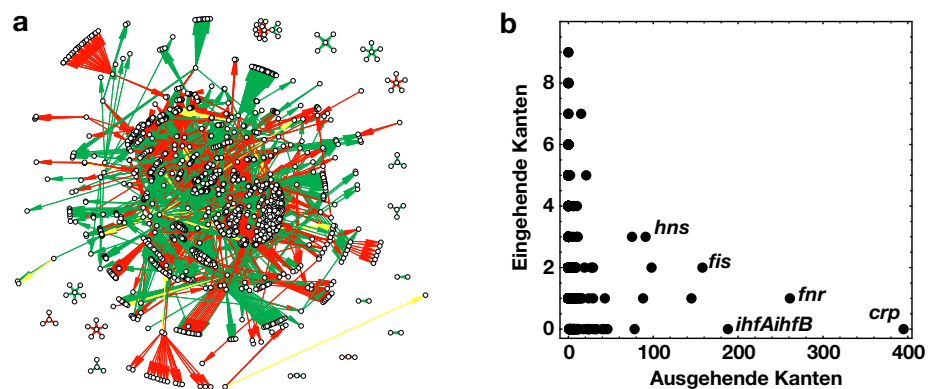


Abb. 4.1. (a) Das Transkriptions-Regulations Netzwerk von *E. coli*, erstellt aus den Daten der RegulonDB. Zu sehen sind 1345 Gene mit 2822 gerichteten Wechselwirkungen. Aktivierung ist als grüner Pfeil, Repression als roter Pfeil, duale Wechselwirkung (in der die tatsächliche Form der Regulierung von weiteren Faktoren abhängt) als gelber Pfeil dargestellt. Der zusammenhängende Subgraph, bestehend aus 1262 Knoten und 2688 Kanten dominiert das Netzwerk. (b) Korrelationsplot zwischen ausgehendem und eingehendem Grad aller Knoten im Netzwerk.

¹ <http://regulondb.ccg.unam.mx/>

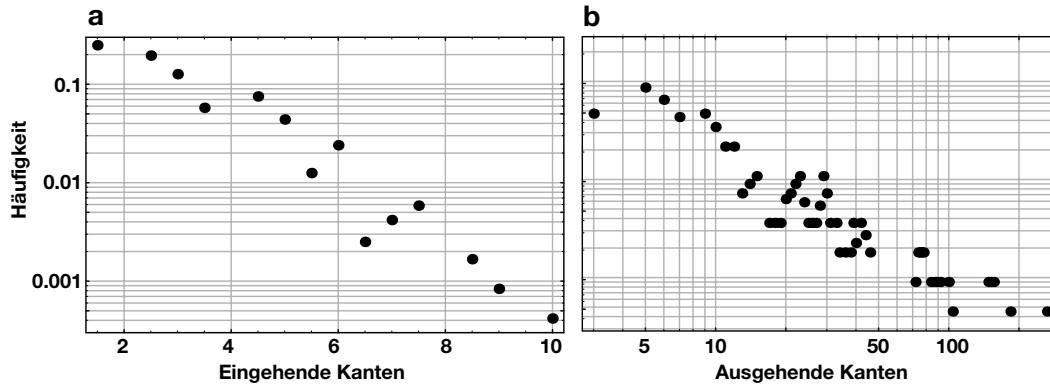


Abb. 4.2. Topologische Eigenschaften der größten zusammenhängenden Komponente des *E. coli* TRN aus Abbildung 4.1. (a) Einfach-logarithmische Darstellung der eingehenden Gradverteilung. (b) Doppelt-logarithmische Darstellung der ausgehenden Gradverteilung.

Art der Regulation (Aktivierung, Inhibition, duale Regulation), die Position der Genen im Genom, etwaige Operonszugehörigkeit sowie die Datenquelle der Wechselwirkung abgerufen werden. Die Rekonstruktion eines Transkriptions-Regulations-Netzwerks (TRNs) aus diesen Daten führt auf einen gerichteten Graphen mit einer dominierenden, 1262 Knoten und 2688 Kanten umfassenden zusammenhängenden Komponente (Abbildung 4.1a). Selbstregulationen sowie mehrfache Kanten zwischen zwei Knoten werden dabei nicht berücksichtigt. Eine Analyse der Gradesequenz des Netzwerks offenbart die Existenz regulierender Hubs (siehe Abbildung 4.1b): Das Gen *crp* reguliert über sein Genprodukt, das Protein CRP, das als TF fungiert, nahezu 395 andere Gene im Netz durch die Bindung des TF an die Promotorregion des regulierten Gens. Weitere Hubs im TRN sind *fnr* (261 regulierte Gene), *ihfAihfB* (188) und *fis* (158). Dabei bezeichnet *ihfAihfB* die Regulation, die durch den dimerischen Transkriptionsfaktor IHF, zusammengesetzt aus den Genprodukten von *ihfA* und *ihfB*, den TFs IhfA und IhfB, vermittelt wird. Der maximale Grad eingehender Kanten im *E. coli* TRN ist 9. Wir betrachten die Verteilung der eingehenden und ausgehenden Kanten in Abbildung 4.2. Die lineare Abhängigkeit der eingehenden Gradverteilung in einfach-logarithmischer Darstellung legt eine Poissonsche Verteilungsfunktion der Form $P(d) \propto e^{-d}$ nahe. Für die ausgehende Gradverteilung deutet der lineare Verlauf in doppelt-logarithmischer Darstellung auf ein Potenzgesetz der Form $P(d) \propto d^{-\gamma}$ hin, eine Eigenschaft, die die skalenfreien Graphen aus Kapitel 1.3 modellieren. Tatsächlich ist eine solche skalenfreie Gradverteilung ein Charakteristikum vieler realer Topologien, von biologischen bis hin zu sozialen und technischen Netzwerken (Albert und Barabási, 2002). Aus einer funktionalen Perspektive sind die Unterschiede zwischen eingehenden und ausgehenden Kanten verständlich: Während einzelne Gene globale regulative Aufgaben übernehmen, um zwischen verschiedenen Zellprogrammen zu schalten, und daher viele andere Gene regulieren, ist die Bindungskapazität einer Promotorregion natürlich beschränkt. Offenbar operiert die Logik der Genregulation also auf einem Netzwerk mit Nachbarschaften begrenzter Größe. Abbildung 4.1b zeigt ferner, dass eingehender und ausgehender Grad antikorreliert sind, globale Regulatoren also nur wenige oder keine Regulatoren besitzen und vielfach regulierte Gene selbst nur wenig oder keine Gene regulieren. Insgesamt entsteht das Bild eines hierarchischen Netzwerks, mit wenigen, regulierenden Knoten in den oberen Ebenen, die Signale an eine Vielzahl von Genen an den Rändern des Netzwerks weiterprozessieren (Ma et al., 2004a; Yu und Gerstein, 2006).

4.2 TRN und Genexpression

Die Miniaturisierung experimenteller Technologien eröffnet die Möglichkeit, die Expression – also die Menge an abgelesenem RNA-Transkript – vieler Gene auf einmal zu messen. Schon mit den ersten

Microarray-Experimenten wurde versucht, solche Expressionsprofile mit den funktionalen Moden einer Zelle in Verbindung zu bringen (Chu et al., 1998; Yoshimoto et al., 2002; Covert et al., 2004; Tu et al., 2005; Marguerat et al., 2006). Kann das TRN – die abstrahierte Form aller bekannten Transkriptions-Regulations Wechselwirkungen – beim Verständnis der funktionalen Abläufe der Zelle – aufgezeichnet in den Microarray Expressionsdaten – dienlich sein? Die Konsistenz zwischen Paaren von Genen und kleinen Motiven (Gruppen von Genen mit bestimmter Wechselwirkungsarchitektur wie zuerst von Shen-Orr et al. (2002) beschrieben) im TRN und entsprechenden Expressionsmustern wurde 2003 in den Artikeln von Gutierrez-Rios et al. (2003) und Herrgard et al. (2003) untersucht. Für Paare von regulierendem Gen und reguliertem Gen ist die Übereinstimmung zwischen der Wechselwirkung im TRN und dem Expressionsverhalten erstaunlich gering: Bei einer aktivierenden Regulation um 15%, bei reprimierenden Regulation sogar nur um 5% (Herrgard et al., 2003) in *E. coli* und um 10% bei *S. cerevisiae*, unabhängig von der Art der Regulation. Die Übereinstimmung zwischen den beiden Datensätzen (auch das TRN ist letztlich nur eine Sammlung experimenteller Einzelbefunden) wird bei mehrere Knoten umfassenden Motiven etwas besser, nie jedoch übersteigt sie 50%. Die Gründe dafür sind vielfältig und reichen von unvollständigen Netzen über Rauschen in Microarray-Daten – insbesondere für reprimierte und daher schwach exprimierte Gene – bis zu der offensichtlichen Vernachlässigung von kooperativen Effekten bei der Bindung mehrerer Transkriptionsfaktoren. Gutierrez-Rios et al. (2003) erarbeiten daher ein ausgefeiltes Protokoll, das Expressionswerte binarisiert und Boolesche Regulationsregeln anwendet, um die Übereinstimmung der beiden Datenstrukturen in *E. coli* zu erhöhen. Tatsächlich erreichen die Autoren auf diese Weise Korrelationen zwischen 50% - 80%. Allerdings stellt eine so aufwendige Interpretationsmaschinerie die Vorteile der einfachen Netzwerkrepräsentation in Frage.

4.2.1 Konsistenz zwischen Expressionsdaten und Paarwechselwirkung im TRN

Wir wollen im folgenden Abschnitt versuchen, TRN-Paarwechselwirkungen in Expressionsdaten wiederzufinden und dabei die Ergebnisse aus Gutierrez-Rios et al. (2003) und Herrgard et al. (2003) mit etwas anderen Mitteln zu reproduzieren (siehe auch Marr et al., 2007c). Dazu wählen wir einen Satz von Microarray-Daten, der über die ASAP-Datenbank² frei verfügbar ist (Glasner et al., 2003) und die Expression eines bestimmten *E. coli* Stamms (K-12 MG1655) unter verschiedenen Umgebungsbedingungen und Mutationen beschreibt. Die insgesamt 41 Experimente aus der *aerobic shift* Reihe (Covert et al., 2004) verwenden Affymetrix-Chips und messen daher absolute Genexpressionswerte. Zu jedem Paar von wechselwirkenden Genen im TRN liefert der ASAP-Datensatz 41 Paare von Expressionswerten. Angenommen Gen *a* hat einen aktivierenden Einfluss auf Gen *b* – durch die Bindung des von Gen *a* synthetisierten TFs –, so erwarten wir eine Korrelation in der entsprechenden Expression: Je mehr Gen *a* transkribiert wird, desto stärker ist sein aktivierender Einfluss auf *b* und entsprechend mehr Transkript von *b* liegt in der Probe vor. Für eine inhibierende Regulation sollten die Expressionswerte gerade antikorreliert sein: Je mehr Transkript von Gen *a* vorliegt, desto stärker wird die Expression von *b* gehemmt, entsprechend weniger Transkript von Gen *b* wird in der Probe zu finden sein. In Abbildung 4.3 wird dieser Ansatz mit den ASAP-Daten überprüft. Abbildung 4.3a zeigt die Korrelation der Expressionswerte von 16 zufällig ausgewählten, aktivierend und reprimierend wechselwirkenden Genpaaren im TRN. Während einige Paare die erwartete Korrelation (grün) bzw. Antikorrelation (rot) aufweisen, sind in anderen Fällen gerade entgegengesetzte, oder keine Korrelationen zu beobachten. Quantifiziert man die Abhängigkeit der Expressionen mit dem Pearsonschen Korrelationskoeffizienten für alle 1215 aktivierenden und 874 reprimierenden regulatorischen Paarwechselwirkungen im TRN, so ergeben sich die in Abbildung 4.3b dargestellten Häufigkeitsverteilungen. Diese sind für aktivierende und reprimierende Kanten im TRN beide um den Nullpunkt zentriert, mit ähnlichen Standardabweichungen. Lediglich eine leichte Erhöhung der reprimierenden Verteilung für negative Korrelationswerte weist auf die unterschiedliche Natur der betrachteten Wechselwirkung hin. Ein ähnlich unspezifisches Bild ergibt sich für die verbleibenden 147 dualen Wechselwirkungen, ebenso für die Ensembles von aktivierenden und reprimierenden Wechselwirkungen, wenn selbstregulatorische Kanten aus dem Netzwerk entfernt wurden.

² <https://asap.ahabs.wisc.edu/asap/ASAP1.htm>

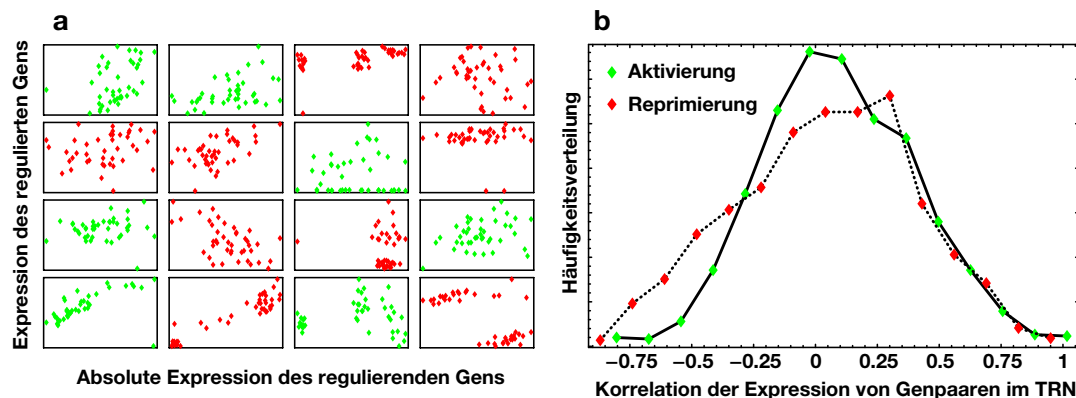


Abb. 4.3. Korrelation der Expressionsdaten von wechselwirkenden Paaren von Genen im TRN. (a) Zufällige Auswahl von Korrelationsplots der Expression eines regulierenden gegen die Expression eines regulierten Gens. Grüne Punkte symbolisieren eine aktivierende Wechselwirkung, rote Punkte Inhibition. (b) Häufigkeitsverteilung der Korrelationskoeffizienten der Expression von Wechselwirkenden Genpaaren im TRN. Es lässt sich kein signifikanter Unterschied der Art der Regulation in den Korrelationen erkennen.

4.2.2 Störungspropagation im *E. coli* TRN

Anstatt wie Herrgard et al. (2003) die Korrelation von Motiven im TRN mit der entsprechenden Expression zu vergleichen wollen wir die Propagation von Störungen untersuchen. Ergibt sich aus dem Übereinanderlegen von Expressionsänderung und Netzwerktopologie – die eine kausale, gerichtete Propagation von Störungen nahelegt – ein konsistentes Bild? Wir betrachten in den ASAP-Daten den Wildtypstamm und vier Mutantenstämme in denen die Gene *arcA*, *fnr*, *oxyR* und *soxS* entfernt wurden. Zu jedem Stamm existieren je drei Microarray-Expressionsdaten unter aeroben und anaeroben Bedingungen. Insgesamt liegen also $5 \cdot 3 \cdot 2 = 30$ Datensätze vor. Als erstes Beispiel betrachten wir die Auswirkung der Gendeletion von *fnr*. Im hierarchischen, gerichteten TRN liegen 444 Gene unterhalb von *fnr*. Damit ist gemeint, dass auf dieses *Sub-*

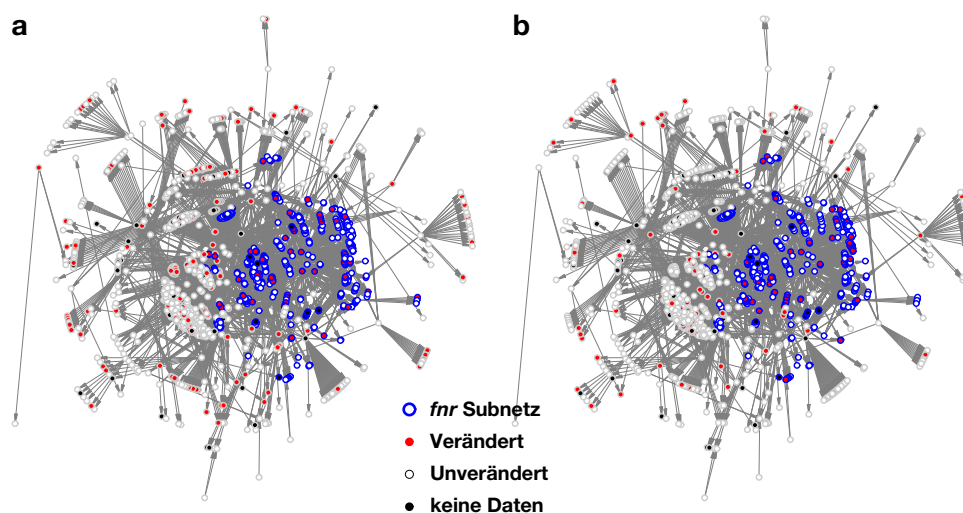


Abb. 4.4. Auswirkung der *fnr* Deletion auf die Expressionswerte der Gene im *E. coli* TRN unter aeroben (a) und anaeroben (b) Bedingungen. Das *fnr* Subnetz ist durch blaue Kreise gekennzeichnet, rote Knoten weisen auf veränderte Expression hin, schwarze Knoten auf Gene ohne Expressionsdaten.

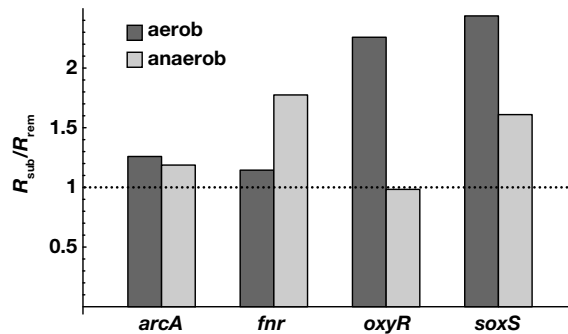


Abb. 4.5. Auswirkung der Gendeletion auf die Expressionswerte der Gene im *E. coli* TRN unter aeroben und anaeroben Bedingungen für vier Mutanten. R_{sub} ist die Anzahl geänderter Expression im Subnetz geteilt durch die Größe des Netzes, R_{rem} entsprechend die relative Anzahl der veränderten Expressionswerte im verbleibenden Netz. R_{sub}/R_{rem} ist ein Maß für den Einfluss des TRN auf die tatsächliche Expressionsänderung.

netz von Genen eine direkte oder indirekte (vermittelt durch andere Gene) regulatorische Wechselwirkung von *fnr* ausgeht. Im gerichteten TRN lässt sich zu jedem dieser Gene ein Weg von *fnr* aus finden. Demnach erwarten wir, dass eine Änderung der Expression von *fnr*, oder im Extremfall, die Deletion des Gens, zu einer deutlichen Änderung der Expression der Gene im Subnetz führt. Gene außerhalb des Subnetzes – im Fall von *fnr* sind das 818 Gene in der größten zusammenhängenden Komponente – sollten dagegen aus der TRN-Perspektive weniger deutlich oder gar nicht von der Mutation betroffen sein. Abbildung 4.4 zeigt die größte zusammenhängende Komponente des *E. coli* TRNs, das darinliegende *fnr* Subnetz (blau) sowie die in ihrer Expression veränderten Gene (rot) für den aeroben (a) und anaeroben (b) Fall. Wir attestieren eine Genveränderung, wenn die Wahrscheinlichkeit, dass die drei Wildtyp und die drei Mutanten Datensätze den gleichen Mittelwert besitzen, unter 5% liegt. Wir verwenden also den *P*-Wert eines zweiseitigen *t*-Tests für die beiden dreikomponentigen Datensätze, um die Signifikanz einer Änderung zu quantifizieren. In Abbildung 4.4 ist deutlich zu erkennen, dass veränderte Expression sowohl innerhalb, also auch außerhalb des Subnetzes der Mutation auftritt.

Um den Effekt weiter zu beschreiben, berechnen wir die relative Anzahl verändert exprimierter Gene im Subnetz, R_{sub} , sowie im verbleibenden Netz, R_{rem} , und teilen diese beiden Größen, R_{sub}/R_{rem} . Wir erhalten so ein Maß für die Kausalität der Änderung im TRN. Abbildung 4.5 zeigt dieses Verhältnis für vier Mutanten im aeroben und anaeroben Fall. Mit einer Ausnahme ist R_{sub}/R_{rem} immer größer 1, d.h. die relative Anzahl von geänderten Genen ist im Subnetz der Mutante immer höher als außerhalb. Der größte Wert, 2.37 ergibt sich für die aerobe *soxS* Deletion, der kleinste, 0.96, für anaerobe *oxyR* Deletion – das Genprodukt von *oxyR* ist gerade für die Antwort der Zelle auf oxidativen Stress verantwortlich (Lynch und Lin, 1996). Werte zwischen 1 und 2 legen also ein Bild nahe, in dem sich die durch die Gendeletion hervorgerufene Änderung des Regulationsapparates vornehmlich in dem Subnetz der Mutante auswirkt. Darüberhinaus ändern sich jedoch auch viele Gene im TRN außerhalb des Subnetzes, also ohne über eine durch das TRN vermittelte, kausale Wechselwirkungen. Um die Regulation jenseits der im TRN vermittelten Wechselwirkungen zu verstehen, müssen weitere Effekte, wie DNA-Supercoiling in Abschnitt 4.3, in Betracht gezogen werden.

4.2.3 Globale Netzwerkeigenschaften in *E. coli* und *S. cerevisiae*

Lässt sich der Einfluss des TRN in anderen Eigenschaften der Expressionsdaten nachweisen? Wir bilden alle verfügbaren Expressionsdaten aus der ASAP-Datenbank auf das TRN ab und vergleichen diese mit topologischen Eigenschaften, insbesondere mit dem Grad, der Gene. In Kapitel 1.3 wurde der Zusammenhang von Grad und dem Eigengewicht eines Knoten, also dem Resultat eines einfachen dynamischen

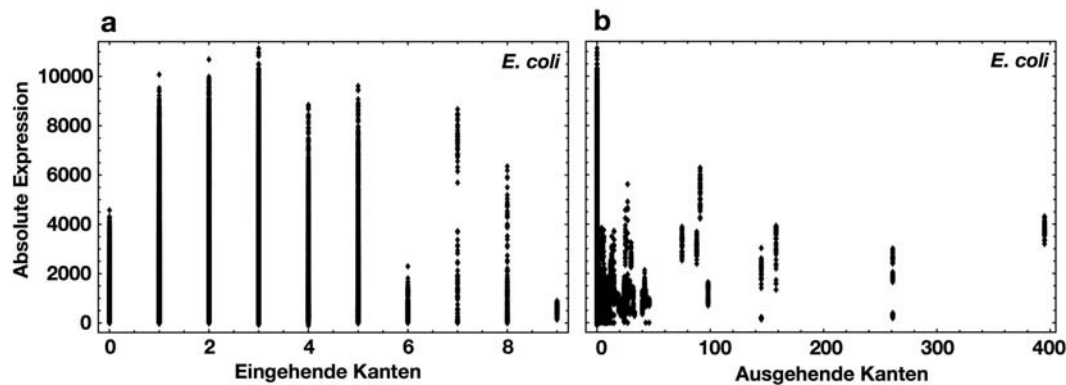


Abb. 4.6. Expressionswerte aller 41 Microarray-Experimente gegen eingehenden (a) und ausgehenden (b) Grad des entsprechenden Knotens im *E. coli* TRN.

Prozesses, in verschiedenen synthetischen Netzen untersucht. Hier suchen wir nach dem Zusammenhang zwischen einfachen topologischen Eigenschaften und der Expression der Gene, also dem Resultat eines komplexen, nichtlinearen Regulationsprozesses. Abbildung 4.6 zeigt die Expressionswerte aller 41 *aero-bic shift* Experimente eines Gens gegen die Zahl der eingehenden (a) und ausgehenden (b) regulatorischen Wechselwirkungen. Während sich für den Grad eingehender Kanten ein recht unspezifisches Bild ergibt, lassen sich Korrelationen der Expression und der Zahl ausgehender Kanten erkennen:

- Die Expression von Genen ohne regulatorischen Effekt variiert deutlich stärker (0 – 12000) als die Expression von Genen mit regulatorischen Wechselwirkung (0 – 6000).
- Die Expression regulatorischer Gene zeigt eine leichte positive Korrelation mit der Anzahl der regulierten Gene.
- Der absolute Wert der Genexpression schwankt nur leicht mit der Veränderung des Sauerstoffgehalts in der Umgebung, sowie mit Mutationen im TRN – offensichtlich ist der regulatorische Grad eines Gens die dominante Komponente.

In Anbetracht der verschwindenden Paarkorrelationen ist die Existenz solcher Korrelationen bemerkenswert. Ein ähnliches, in manchen Eigenschaften sogar noch deutlicheres Bild ergibt sich in Abbildung 4.7 für ein Operon-TRN (Shen-Orr et al., 2002). Dabei werden Gene mit einer gemeinsamen Promotorregion

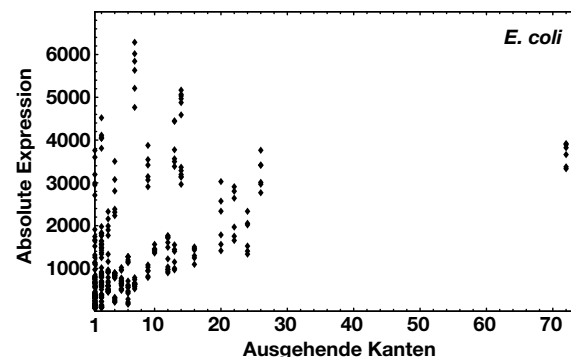


Abb. 4.7. Expressionswerte aller 41 Microarray-Experimente gegen ausgehenden Grad des entsprechenden Knotens im *E. coli* Operon TRN.

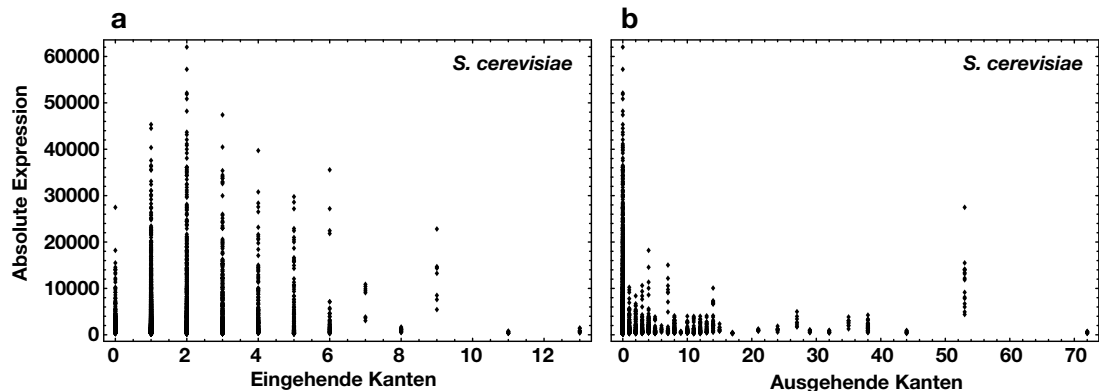


Abb. 4.8. Expressionswerte von 8 Microarray-Experimente gegen eingehenden (a) und ausgehenden (b) Grad des entsprechenden Knotens im *S. cerevisiae* TRN mit 664 Knoten und 1078 Kanten.

zu einer Einheit, einem *Operon* zusammengefasst. Die Daten des entsprechenden Netzwerks, generiert auf Basis der RegulonDB Wechselwirkungen, sind frei zugänglich auf den Seiten des Weizmann Instituts³. In Abbildung 4.7 scheinen die regulatorischen Operons in zwei Gruppen zu zerfallen: Eine stärker exprimierte, und eine schwächer exprimierte Gruppe. Es lässt sich jedoch keine klare funktionale Abgrenzung der beiden Gengruppen, beispielsweise beim Abbilden von *Gene-ontology*-Klassen, finden.

Wir wollen überprüfen, ob sich in den Expressionseigenschaften der *E. coli* TRN Gene allgemeine Prinzipien, oder lediglich speziesspezifische Artefakte wiederfinden lassen. Daher korrelieren wir Expressionsdaten mit topologische Observablen des entsprechenden TRNs für *S. cerevisiae* – das zweite derzeit verfügbar TRN neben *E. coli*. Verschiedene Repräsentationen des Hefe-Netzwerks existieren in der Literatur, abhängig von der Methode und der Datengrundlage (Lee et al., 2002; Milo et al., 2002). Wir betrachten im folgenden das Netzwerk, das auf Daten der Yeast Proteome Database (Costanzo et al., 2001) basiert, von der Gruppe um Uri Alon zusammengestellt und kuratiert wird und frei zugänglich ist⁴ (Andere Repräsentationen, wie die von Guelzim et al. (2002) oder Luscombe et al. (2004), führen auf ähnliche Ergebnisse). Es umfasst 664 Knoten und 1078 gerichtete Kanten. Wir verwenden Expressionsdaten aus Microarray-Experimenten von Yoshimoto et al. (2002), die über die ArrayExpress-Datenbank frei zugänglich sind⁵. In dieser Versuchsreihe wurde die Expressionsänderung der Hefegene durch Zugabe von Ca^{2+} - und Na^{+} -Ionen zeitaufgelöst untersucht. Abbildung 4.8 zeigt die zu 4.6 analogen Ergebnisse für Hefe, wobei alle 8 Datensätze auf den Grad des entsprechenden Gens abgebildet wurden. Wie schon bei *E. coli* zeigen sich auch hier Muster in der Darstellung der Expression gegen die Zahl der regulierten Gene: Gene ohne regulatorische Wechselwirkung weisen ein sehr viel weiteres Spektrum an Expressionswerten auf als regulierende Gene, die Änderung der Expression gegenüber wechselnden Umweltbedingungen – in diesem Fall die Änderung der Ca^{2+} und Na^{+} Konzentrationen – ist wie bei *E. coli* moderat. In Abbildung 4.8(a) ist die absolute Genexpression gegen die Zahl regulierender Gene aufgetragen. Bei Hefe ist das Maximum von Regulatoren für ein Gen $d = 13$ – dabei zeigt sich ein leichtes Abfallen des Expressionsspektrums mit dem Grad eingehender Kanten, mit einem Maximum bei $d = 2$.

Lassen sich die beobachteten Eigenschaften der Expressionsmuster als Signaturen von dynamischen Prozessen auf dem TRN verstehen? In Abschnitt 1.2 hatten wir mit der Potenzmethode zur Berechnung des PF-Eigenvektors ein einfaches Beispiel für eine lineare Anregungsdynamik kennengelernt. Die Übertragung einer so einfachen dynamischen Sonde von ungerichteten auf gerichtete Graphen wurde am Ende des gleichen Abschnitts besprochen: um Knoten ohne eingehende Kanten mit einem von Null verschiedenen Eigengewicht zu belegen, wurde von Hubbel (1965) die Einführung eines konstanten Eigengewichts

³ www.weizmann.ac.il/mcb/UriAlon/

⁴ www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/yeastData.mat

⁵ www.ebi.ac.uk/arrayexpress/, Experimentnummer E-SMDB-7

vorgeschlagen. Das Konzept der *Hubs-and-authorities* von Kleinberg (1999) kommt ohne einen zusätzlichen Parameter aus, indem es jedem Knoten zwei nichtnegative Gewichte x_i und y_i zuordnet. Im Sinne des TRN messen diese beiden Größen die Wichtigkeit des Knotens als regulierte Autorität – das heißt als multireguliertes Gen, dessen Zustand viele andere Regulatorzustände im System in einer prozessiert Form inkorporiert –, bzw. als regulierender Hub im System. Eine rekursive Anwendung der Gleichung (1.6) führt indes auf eine einfache Bestimmungsgleichungen der beiden Vektoren. So ist das Autoritäts-Gewicht \bar{x} der PF-Eigenvektor zur Matrix $A^T A$, das Hub-Gewicht \bar{y} der PF-Eigenvektor zur Matrix $A A^T$. Beide Größen lassen sich leicht für das Hefe-TRN berechnen und gegen die Grade der Knoten auftragen. Abbildung 4.9 zeigt das Ergebnis der Potenzmethode (siehe Gl. (1.5)): Offensichtlich prägt die komplexe TRN-Struktur dieser dynamischen Sonde ein Muster auf, das die trivialen Zusammenhänge verschwinden lässt. So ist nur eine kleine Korrelation des Hub-Gewichts mit der Zahl ausgehender Kanten, und keinerlei Korrelation des Autoritäts-Gewicht mit der Zahl eingehender Kanten erkennbar. Darüberhinaus ist es interessant zu sehen, wie letztere Größe einige charakteristische Muster aus Abbildung 4.8 reproduziert: Ein großes Spektrum an Werten für kleine eingehende bzw. keine ausgehenden Grade, sowie beschränkte Anregungswerte für Knoten mit hohem Grad. Der dem Autoritäts-Gewicht zugrundeliegende Prozess bei der Bestimmung des PF-Eigenvektors ist dabei die Potenzmethode auf einem ungerichteten Graphen mit der Adjazenzmatrix $A^T A$. Dieser, für die Berechnung des Autoritäts-Gewichts effektive Graph besitzt im Fall des Hefe-TRN 664 Knoten und – nach Entfernung von multiplen Kanten und Selbstregulation – 10188 Kanten. Der hohe Vernetzungsgrad kommt dabei durch verknüpfte Knoten zustande, die im Originalgraphen durch einen Weg mit zwei entgegengesetzt gerichteten Kanten verbunden sind.

Will man die Ähnlichkeit von Abbildung 4.8 und 4.9 aus dieser Perspektive interpretieren, so muss man einen dynamischen Effekt entgegen der Regulationsrichtung postulieren. Dabei darf nicht vergessen werden, dass diese Ähnlichkeit vor allem eine qualitative ist: Die tatsächliche Korrelation zwischen Expressionswerten und Autoritätsgewicht ist verschwindend gering (0.01). Es wird mit dieser Sonde also nur ein Aspekt des dynamischen Potentials der TRN-Topologie beschrieben.

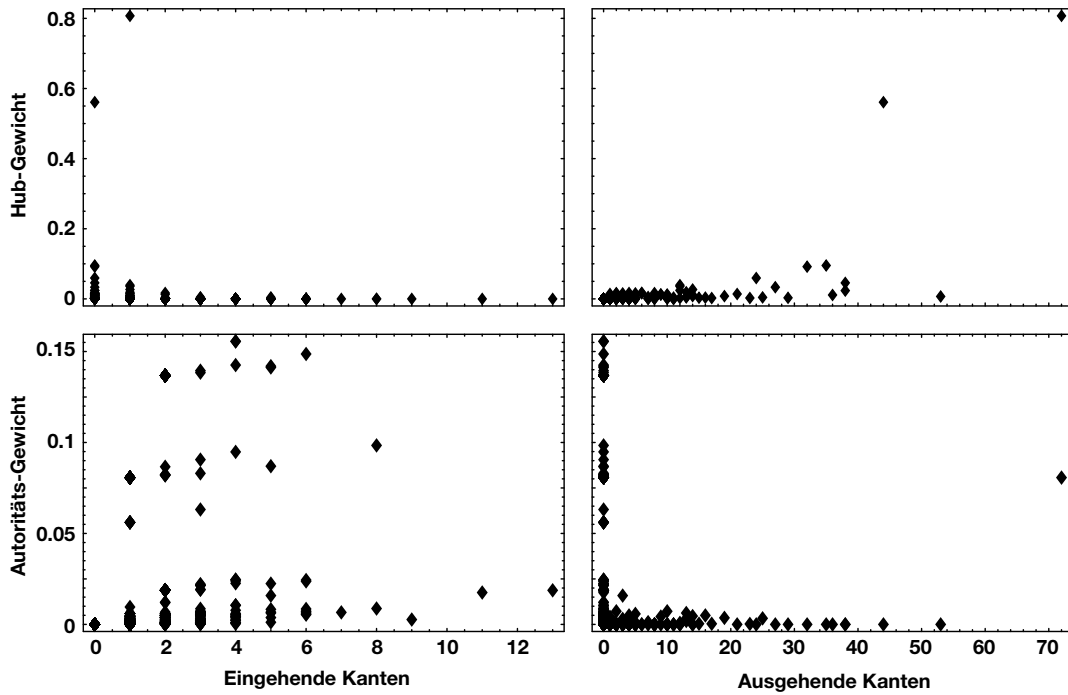


Abb. 4.9. Die *Hubs-and-Authorities*-Gewichtung (Kleinberg, 1999) auf dem Hefe-TRN gegen die Zahl der eingehenden und ausgehenden Kanten.

4.3 Kontrollmechanismen der Transkriptions-Regulation in *E. coli*

Im vorhergehenden Abschnitt wurde gezeigt, dass Teilaspekte der gemessenen Expressionsmuster in *E. coli* mit Eigenschaften des TRN erklärt werden können. Daneben wurde aber auch deutlich, dass das TRN alleine nicht ausreicht, um Veränderungen der Expressionsmuster, beispielsweise aufgrund von Mutationen, vollständig zu erklären. Ein regulatorischer Effekt, der nicht auf der Wechselwirkung zwischen Transkriptionsfaktoren und deren Bindestellen aufbaut, sondern sich auf benachbarte Gene in der DNA bezieht, ist das sogenannte *Supercoiling* (Travers und Muskhelishvili, 2005b,a). Damit ist die negative Superhelizität des Chromosoms gemeint, das Verdrillen der Doppelhelix zu Keulen. Auf den ersten Blick hat das Aufwickeln der DNA zu Supercoils praktische Gründe: Um das 4.6 Megabasenpaare (MBp) lange, zirkuläre Genom in das Zellvolumen des Bakteriums einzupassen, muss es etwa 1000-fach komprimiert werden (Travers und Muskhelishvili, 2005a). Daneben wirkt sich die Strukturänderung aber auch auf den Expressionsgrad der beteiligten Gene aus: Je nach Stärke der Verdrehung der DNA wird das Binden der RNA-Polymerase – ein Makromolekül, das die Transkription, das Abschreiben der Gen-Nukleotidsequenz in einsträngige mRNA Kopien, leistet – erleichtert oder behindert. Mit lokalen Strukturänderungen konnten Postow et al. (2004) zeigen, dass die Domänengröße der einzelnen Coils um 10 kbp liegt. Dieser Wert liegt interessanterweise in der Größenordnung der elektronenmikroskopischen Messungen von Supercoils in isolierten *E. coli* Chromosomen (Postow et al., 2004), was eine Übereinstimmung der funktional und visuell ermittelten DNA-Strukturen nahelegt (Travers und Muskhelishvili, 2005a).

Um den regulatorischen Effekt von Supercoiling mit der durch Transkriptionsfaktoren vermittelten Regulation zu vergleichen, analysieren wir einen Datensatz mit komparativen Microarray-Daten (Blot et al., 2006), in dem sowohl die negative Superhelizität σ , als auch das TRN verändert werden (siehe Abbildung 4.10). Drei *E. coli* Stämme – Wildtyp, *fis*-Mutante, *hns*-Mutante – werden unter relaxierter und erhöhter negativer Superhelizität untereinander und miteinander verglichen. Die zu analysierenden Daten gruppieren sich also in drei Intrastamm-Experimente i_1, i_2, i_3 – in denen der Effekt der Superhelizität auf die Genexpression in jedem Stamm separat verglichen wird – und vier Interstamm Experimente I_1, I_2, I_3, I_4 – in denen der Wildtyp mit jeder Mutante für beide Supercoiling-Bedingungen verglichen wird. Jedes Experiment besteht dabei aus vier Microarray Messungen mit je zwei biologischen und zwei technischen Replikaten. Ein Gen wird dann als differentiell exprimiert bezeichnet, wenn der *P*-Wert eines einseitigen *t*-Tests der beiden Intensitätsdatensätze kleiner als 0.05 ist (siehe Blot et al., 2006, für weitere methodische Details). In den Intrastamm-Experimenten stimmen die genetische Information, und damit das TRN der beiden untersuchten

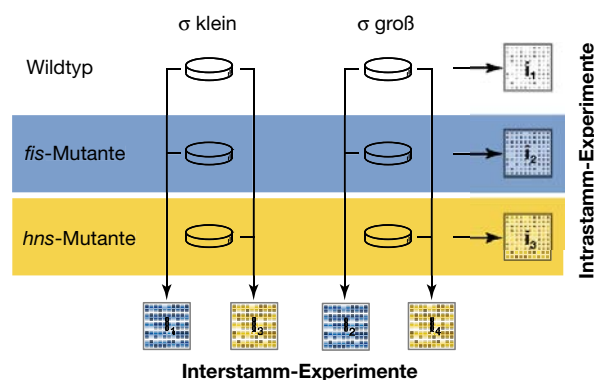


Abb. 4.10. Nomenklatur und Vergleichsstruktur der analysierten Daten. Aus Wildtyp, *fis*-Mutante und *hns*-Mutante in jeweils relaxierter und erhöhter negativer Superhelizität σ lassen sich drei Intrastamm-Datensätze sowie vier Interstamm-Datensätze generieren. Bei ersteren wird die analoge Komponente der Kontrolle variiert, während der genetische Hintergrund, die digitale Komponente, konstant bleibt. Bei letzteren werden verschiedene Stämme verglichen, d.h. die digitale Komponente wird variiert, während die analoge Komponente, die Stärke der DNA-Superhelizität, konstant bleibt.

Zelltypen überein. Differenziell exprimierte Gene geben in diesem Fall Aufschluss über die Art und Weise, wie die verschiedenen Stämme mit verändertem Supercoiling umgehen. In den Interstamm-Experimenten dagegen werden verschiedene Zelltypen mit unterschiedlichem Genom und TRN, aber unter der in gleicher Weise veränderten Superhelizität miteinander verglichen. So lassen sich die Reaktionen der Stämme auf geändertes Supercoiling miteinander vergleichen. Abbildung 4.10 zeigt die verschiedenen Stämme und das Zustandekommen der sieben Datensätze schematisch.

4.3.1 Digitale Kontrolle

Wie verändern sich die Expressionsmuster, wenn sich Wildtyp- und Mutantenzellen an verschieden starke Verdrillungen der DNA anpassen müssen? Inwieweit hat das TRN auf eine solche Adaption Einfluss? Wir nähern uns dieser Fragestellung, indem wir aus dem Satz differentiell exprimierter Gene ein effektives TRN konstruieren. Die spezifische Regulation eines Gens durch das Binden eines TFs an die entsprechenden Promotorregion bezeichnen wir als *digital*. Das komplette, alle bislang experimentell verifizierten Wechselwirkungen umfassende TRN der RegulonDB stellt den Möglichkeitsraum der digital vermittelten Regulation dar. Dass bei wechselnden Umweltbedingungen und funktionalen Moden nur bestimmte Teile des kompletten TRN verwendet werden, zeigen Luscombe et al. (2004) für Hefe. Die Autoren erzeugen dazu effektive TRNs aufgrund von absoluten Expressionsdaten verschiedener Zellzustände, analysieren die entstehenden Netze und finden 'endogene' und 'exogene' Strukturen: komplexe Netze mit langen Weglängen, wenigen Hubs und starker Clusterung (endogen) für die Zellprogramme bei Zellzyklus und Sporenbildung, dagegen von Hubs dominierte, einfache Topologien mit kurzen Wegen und schwacher Vernetzung (exogen) für Zellen unter oxidativem Shift, DNA Beschädigung und wechselnden Umweltbedingungen (Stress). Im vorliegenden Fall erzeugen wir ein effektives TRN, indem wir pro Experiment alle differentiell exprimierten Gene auf das TRN abbilden und alle nicht-differentiell exprimierten Gene aus dem Netzwerk löschen.

Um den Einfluss der digitalen Kontrolle, vermittelt durch das TRN, auf das System abzuschätzen, berechnen wir Kontrollkoeffizienten R . Er ist als die Zahl aller verbundenen Knoten im effektiven TRN geteilt durch die Zahl der isolierten Knoten ohne jede Wechselwirkung definiert,

$$R = \frac{N_{d>0}^{\text{eff}}}{N_{d=0}^{\text{eff}}} . \quad (4.1)$$

Abbildung 4.11a zeigt R für die effektiven TRNs aller sieben Experimente. Dabei heben sich die Intrastamm-Experimente (i) mit R Werten über 0.5 klar von den Intrastamm-Experimenten ab. Diese Diskrepanz

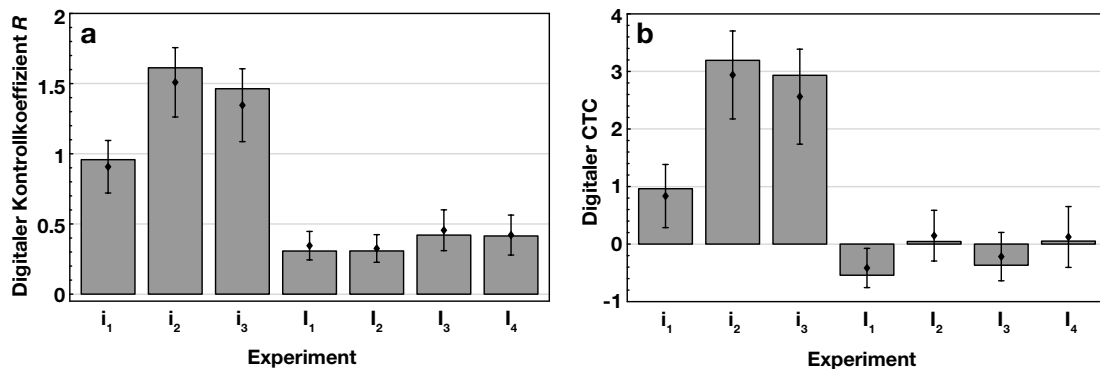


Abb. 4.11. Kontrollkoeffizient R (a) und zugehöriger z -score, der digitale CTC (b) für die effektiven TRNs aller sieben Experimente. Die Intrastamm-Daten der beiden Mutanten, i_2 und i_3 , heben sich sowohl in ihren absoluten Werten (R), als auch in ihren z -scores von den übrigen Experimente deutlich ab. Die Fehlerbalken zeigen die Streuung der Werte, wenn 10% der Daten mit zufällig ausgewählten Genen ersetzt werden.

wird verständlich, wenn wir die zugrundeliegenden Daten in Betracht ziehen: Die differentiell exprimierten Gene der *i*-Reihe werden durch das TRN des entsprechenden Stamms kontrolliert, Wechselwirkungen zwischen den sich gegenseitig regulierenden Knoten bilden sich also in den Expressionsmustern ab. Die Daten der *I*-Reihe dagegen entstehen aus einer Überlagerung der TRNs der beiden verglichenen Stämme – dementsprechend geringere Konsistenzen spiegeln sich in den entsprechenden Expressionsdaten wider.

Wir berechnen die topologische Observable R hier für Netze unterschiedlicher Größe, da die Anzahl der differentiell exprimierten Gene sowie die Anzahl der Knoten N der effektiven TRNs von Experiment zu Experiment variiert. Der Kontrollkoeffizient R hängt jedoch stark von der Knotenzahl ab: Ist $N = 1$, so existieren im effektiven TRN keine Kanten und $R = 0$; dagegen geht R gegen unendlich, wenn N gegen die Netzwerkgröße des kompletten TRN tendiert. Die absoluten R -Werte aus Abbildung 4.11a müssen also statistisch untersucht werden. Wir vergleichen dazu die tatsächlichen Daten mit einem Nullmodell, wobei wir für jedes Experiment effektive TRNs konstruieren, die auf der gleichen Zahl von Genen, aber einer zufälligen Auswahl basieren. Abbildung 4.12 zeigt die induzierten effektiven TRNs, basierend auf den tatsächlichen Daten und dem adäquaten Nullmodell. Eine deutlich stärkere Vernetzung der beiden Mutanten-Interstamm-Experimente i_2 und i_3 gegenüber dem entsprechenden Nullmodell ist klar zu erkennen.

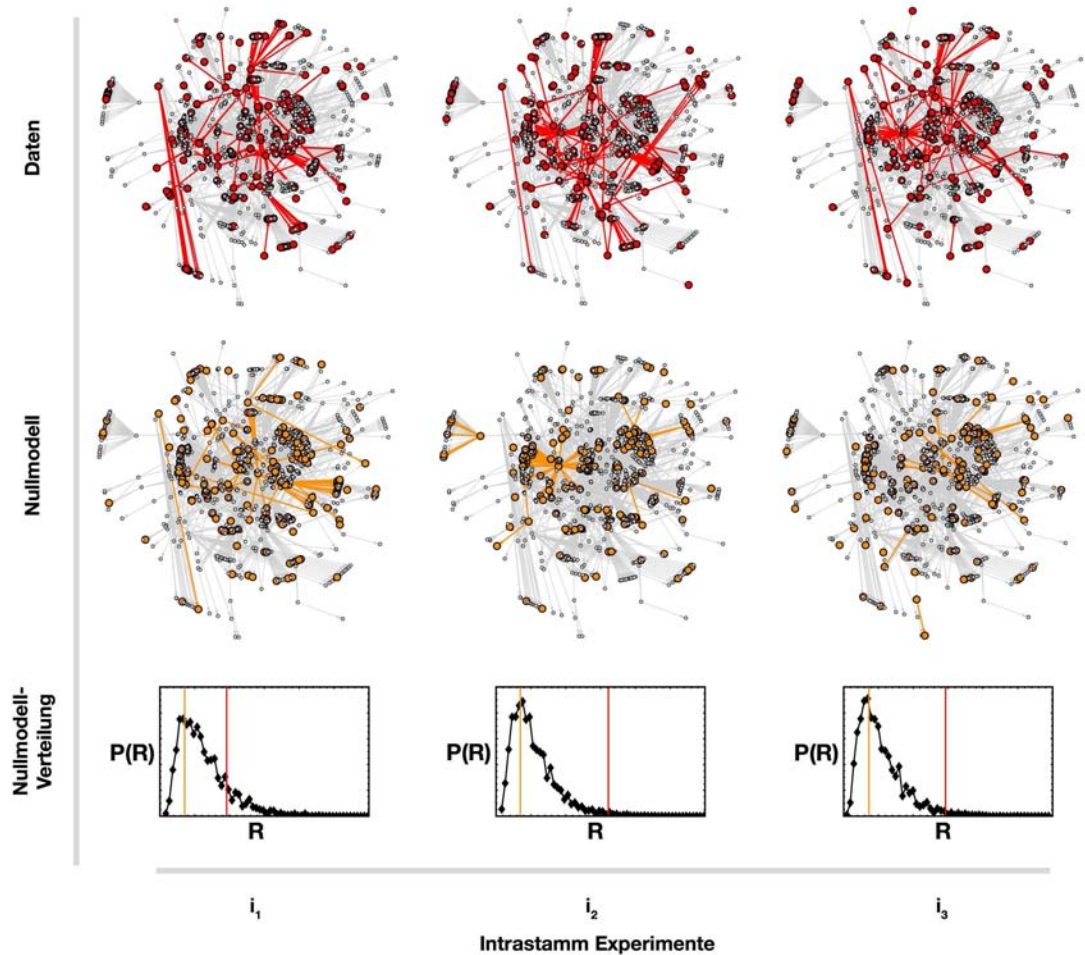


Abb. 4.12. Vergleich der effektiven TRNs, induziert durch die tatsächlichen Expressionsdaten der Intra-Stamm-Experimente (oben) mit dem entsprechenden Nullmodell (Mitte), generiert aus der gleichen Anzahl von zufällig ausgewählten Genen. Darunter ist die Häufigkeitsverteilungen der R -Werte der Nullmodelle dargestellt.

Die Abweichung vom Nullmodell effektiver TRNs verschiedener Größe lässt mit einem sogenannten *z-score* quantifizieren. Dazu berechnet man die Abweichung des tatsächlichen Wertes x vom Mittelwert der Nullmodelle \bar{n} in Einheiten der Standardabweichung σ_n der Nullmodelle, $z = (x - \bar{n})/\sigma_n$. Abbildung 4.11b zeigt die zu R gehörenden *z-scores* aller Experimente. Wir bezeichnen diese Größe als Kontrolltyp-Signifikanz (engl. *control type confidence*, kurz CTC). Die Unspezifität der Interstamm-Experimente bezüglich des benutzten TRNs durch die Überlagerung zweier Stämme manifestiert sich hier in $|\text{CTC}| \leq 0.7$. In der *i*-Reihe wird die in Abbildung 4.11a angedeutete Trennung von Wildtyp und Mutante noch deutlicher: i_2 und i_3 zeigen beide CTCs weit über 2, während beim Wildtyp ein $\text{CTC} < 1$ festzustellen ist (siehe dazu auch die Häufigkeitsverteilungen der R -Werte im Nullmodell in Abbildung 4.12).

Die Analyse zeigt, dass sich die durch das TRN vermittelte Kontrolle in der Anpassung der Bakterien an veränderte Superhelizität zwischen Wildtyp und Mutanten stark unterscheidet. Während der Wildtyp nicht signifikant vom Nullmodell abweicht, zeigen die Mutantenstämme eine erhöhte Kontrolle durch das TRN im System. Das ist gerade deshalb überraschend, weil die beiden entfernten Gene, *fis* und *hns*, im TRN die Rolle von Hubs übernehmen, also viele andere Gene regulieren. Intuitiv würde man erwarten, dass ein Löschen von Hubs im TRN zu einer Schwächung dieses Regulationssystems führt. Offensichtlich geht die Zelle jedoch anders mit dem Verlust eines globalen Regulators um: Die Kontrolle des TRN auf die Regulation der übrigen Gene wird im Vergleich zum Wildtyp deutlich erhöht.

4.3.2 Analoge Kontrolle

Im vorhergehenden Abschnitt wurde zur Beschreibung der spezifische Regulation durch TFs der Begriff der digitalen Kontrolle eingeführt. Entsprechend wollen wir nun die Regulation vieler, räumlich benachbarter Gene durch die kontinuierliche Verdrehung der DNA-Doppelhelix zu Supercoils als *analog* bezeichnen.

Um den Anteil der analogen Kontrolle, vermittelt durch DNA Supercoils, in Wildtyp und Mutante quantitativ zu bestimmen, bedienen wir uns der gleichen Werkzeuge wie in Abschnitt 4.3.1 und den Daten der vier Interstamm-Experimente $I_1 - I_4$. Dabei werten wir die komparativen Microarray-Daten je nach Typ der Mutante unterschiedlich aus: Bei den Interstamm-Experimenten der *fis*-Mutante werden Expressionswerte mit positiver Log-ratio (d.h. erhöhter Expression im Wildtyp-Experiment) dem Wildtyp, mit negativer Log-ratio der Mutante zugeordnet, da das Binden von FIS aktivierend auf die benachbarten Gene wirkt (Travers und Muskhelishvili, 2005b; Blot et al., 2006). Die durch FIS etablierten, superhelikalen Loops erleichtern das Binden der Polymerase und sind darüberhinaus sehr flexibel. H-NS dagegen wirkt als universeller Repressor (Dorman, 2004), die so erzeugten Strukturen sind starr und verhindern das Binden der Polymerase. Im Fall der *hns*-Mutante schlagen wir daher Expressionswerte mit positiver Log-ratio der Mutante, solche mit negativer Log-ratio dem Wildtyp zu. Das Vorhandensein der Strukturproteine ändert jedoch nicht die Superhelizität, vielmehr lokalisieren und stabilisieren FIS und H-NS die vorhandenen Supercoils, allerdings in sehr unterschiedlicher Weise. Wir erhalten so pro Experiment zwei Listen mit den Genen, die im Vergleich zum jeweils anderen Stamm stärker exprimiert sind. Aus jeder Liste konstruieren wir nun ein Nachbarschaftsnetzwerk (*gene proximity network*, kurz GPN), in dem zwei Gene genau dann verknüpft sind, wenn ihre ORFs auf dem Genom weniger als t Basenpaare auseinander liegen. Mit dem Schwellenparameter t lässt sich der Vernetzungsgrad des entstehenden GPN durchstimmen: Für $t = 0$ existieren keine Kanten im GPN, erreicht t die Länge des Genoms, so ist das GPN vollständig vernetzt. Bei der eben erklärten Zuweisung bleiben freilich die Gene unberücksichtigt, die in beiden Stämmen gleichermaßen exprimiert werden, da sie in den komparativen Datensätzen mit einer Log-ratio von 0 auftauchen. Dieser Fehler tritt jedoch für beide Stämme in gleicher Weise auf und beeinträchtigt die Analyse dann nicht, wenn sie Observablen verwendet, die unabhängig von der Netzwerkgröße sind.

Wie schon bei der Analyse der digitalen Kontrolle, vermittelt durch das TRN, wählen wir auch hier zunächst den in Gleichung (4.1) definierten Kontrollkoeffizienten R als Observable. R lässt sich für das gerichtete TRN und das ungerichtete GPN in gleicher Weise als das Verhältnis von vernetzten zu isolierten Knoten berechnen. Da wir auch hier Netze verschiedener Größe vergleichen (im Experiment I_1 werden 230 Gene

dem Wildtyp, 330 Gene der *fis* Mutante zugeordnet) können wir den Anteil analoger Regulation nicht mittels der absoluten Werte des Kontrollkoeffizienten vergleichen. Vielmehr erzeugen wir wiederum für jedes Netz ein Nullmodell mit gleich vielen Genen, positionieren diese zufällig auf dem Genom und konstruieren ein GPN mit dem selben Schwellenparameter t . Auf diese Weise lassen sich Wildtyp und Mutante in Bezug auf die räumliche Ordnung der transkribierten Gene vergleichen. Abbildung 4.13 zeigt schematisch die Vorgehensweise bei der Analyse der analogen Kontrolle, sowie das GPN für die Expressionsdaten aus I_1 mit einem Schwellenparameter $t = 5000$ (motiviert durch die in Postow et al. (2004) postulierten typische Länge der Supercoils von 10kbp) und ein Nullmodell. Quantitativ fassen wir die Abweichung des Kontrollkoeffizienten vom Nullmodell wie schon in 4.3.1 durch die Berechnung des entsprechenden z -scores, den wir als analogen CTC bezeichnen.

Wir analysieren alle Interstamm-Experimente mit dem eben beschriebenen Protokoll und trennen die Expressionsdaten nach ihrer Log-ratio in \pm auf. Abbildung 4.14a zeigt den analogen CTC der GPNs der jeweiligen Expressionsdaten mit einem Schwellenwert von $t = 5000$. Für die beiden Interstamm-Experimente, die den Wildtyp mit der *fis*-Mutante vergleichen (I_1 und I_2), zeigen sich erhöhte CTCs für die Expressionswerte mit einer positiven Log-ratio (+). Im Fall der *hms*-Mutante dagegen (I_3 und I_4) ergeben sich erhöhte CTCs für die Gene mit negativer Log-ratio. Dieser Gegensatz ist unabhängig vom Schwellenparameter t . Abbildung 4.14b zeigt die Differenz der CTCs positiver und negativer Log-ratios aller vier Experimente für $t = 5000$. Abbildung 4.14c zeigt die Verläufe der Differenzen unter Variation des Schwellenparameters t . Die Differenzen ΔCTC der Mutanten unterscheiden sich über den gesamten sinnvollen Wertebereich von t . Zusammen mit der Information über die aktivierende Wirkung von FIS und die re-

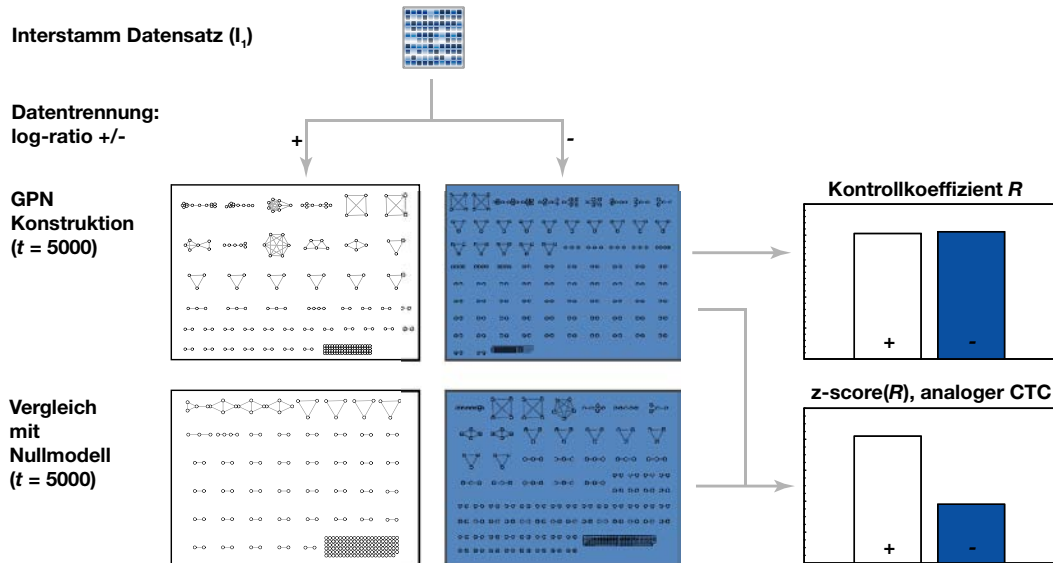


Abb. 4.13. Schema der Datenanalyse der Interstamm-Experimente am Beispiel I_1 . Aufgrund der aktivierenden Regulation von *fis* werden Gene mit einer positiven Log-ratio (+) dem Wildtyp, solche mit einer negativen Log-ratio (-) der Mutante zugeordnet. Die Fixierung des Schwellenparameters $t = 5000$ erlaubt die Konstruktion zweier GPNs und die Berechnung des Kontrollkoeffizienten R , getrennt für Wildtyp und Mutante, der den Einfluss des Supercoilings im jeweiligen Stamm reflektiert. Zu jedem GPN werden außerdem Nullmodelle erzeugt, mit der gleichen Anzahl von Genen, aber zufällig auf das Genom verteilt. Aus dem Mittelwert und der Standardabweichung der Kontrollkoeffizienten der Nullmodelle wird dann ein z -score, der CTC, berechnet. Das Wildtyp-GPN ist im Vergleich zum Nullmodell deutlich höher vernetzt. Das GPN der *fis*-Mutante umfasst mehr Knoten als das Wildtyp-GPN, zeigt einen ähnlichen Kontrollkoeffizienten, weicht vom Nullmodell aber weniger stark ab. Die Kontrolle der Gene durch analoge Regulation ist im Wildtyp also deutlich stärker.

primierenden Wirkung von H-NS lässt sich Abbildung 4.14 interpretieren: Die Wildtyp-Gene zeigen im Vergleich mit den Genen der Mutante einen deutlich erhöhten z -score des Kontrollkoeffizienten R . Die Kontrolle durch das GPN ist im Wildtyp also deutlich stärker. Dieses Verhältnis ist nicht trivial, da die strukturstabilisierenden Proteine FIS und H-NS in den Mutanten zwar fehlen, die Größe der Superhelizität aber in allen verglichenen Stämmen gleich ist. Die erniedrigte, durch das GPN vermittelte Regulation in den Mutanten weist also auf den Verlust der analogen Kontrollinstanz in diesen Stämmen hin.

Die Existenz von Operons, also Gruppen von Genen, die über die gleichen Promotoren, und damit konzentriert reguliert werden, beeinflusst die Konstruktion von GPNs, da diese Gene einerseits in gleicher Weise auf Regulation reagieren, andererseits auf dem Genom räumlich sehr nahe beieinander liegen. Würden im Wildtyp besonders viele Operons exprimiert, so ergäbe unsere obige Analyse trivialerweise stark geclusterte Wildtyp-GPNs mit hohen Kontrollkoeffizienten. Gerade die z -scores wären für Operon-dominierte Expression sehr hoch, da eine zufällige Verteilung von Genen auf nicht-operonartige Strukturen führt. Wir wiederholen daher unsere Analyse, indem wir die entsprechenden Expressionswerte den zugehörigen Operons, und nicht länger den Genen zuweisen. Dabei werten ein Operon dann als vom Supercoiling betroffen, sobald eines der Gene im Operon positive bzw. negative Log-ratio zeigt. Zusammen mit dem Schwellenparameter t , der zwei Operons dann mit einer Kante verknüpft, wenn ihr Abstand auf dem Genom kleiner als t ist, lässt sich analog zum GPN ein Operon-Nachbarschaftsnetzwerk erstellen. Die analogen CTCs dieser Netzwerke zeigen die gleiche Tendenz wie Abbildung 4.14: Eine verstärkte analoge Kontrolle der Wildtyp-Regulation im Vergleich zu den beiden Mutantenstämmen.

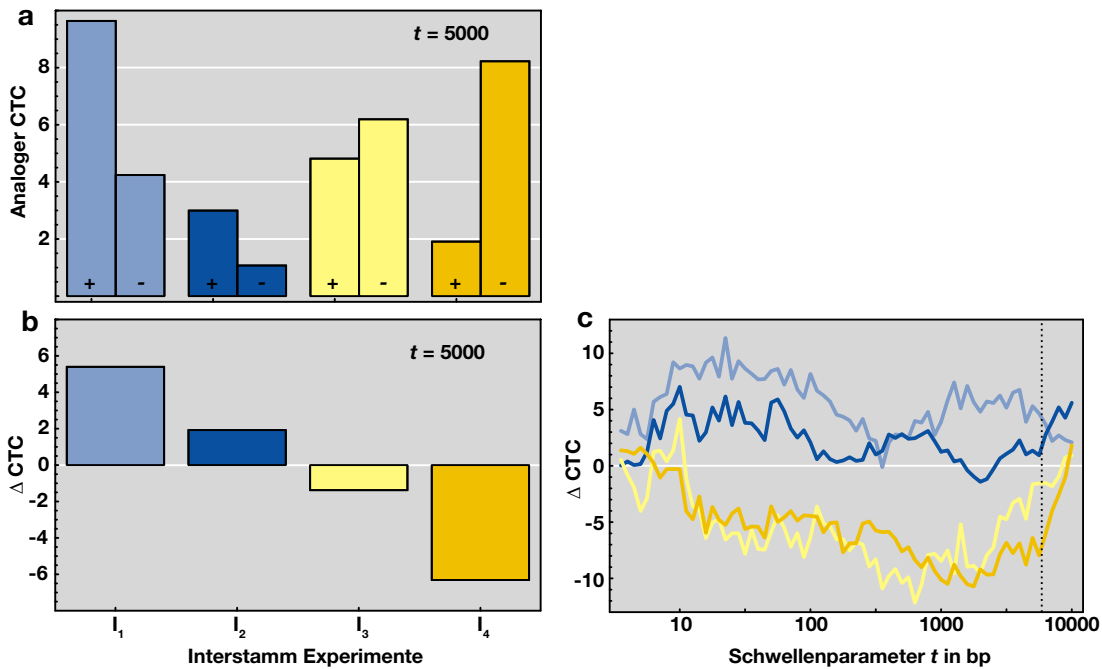


Abb. 4.14. Analoge Kontrolle. (a) Analoges CTC für Daten mit positiver (+) und negativer (-) Log-ratio der vier Interstamm-Experimente mit einem Schwellenparameter $t = 5000$. (b) Die Differenzen der analogen CTCs mit positiver bzw. negativer Log-ratio aus (a). Die *fis*-Daten (blau) zeigen eine erhöhte analoge Kontrolle für Gene mit positiver Log-ratio, die *hns*-Daten (gelb) eine erniedrigte Kontrolle für Gene mit negativer Log-ratio. Bezieht man das Wissen über die aktivierende Wirkung von FIS, und die inhibierende Wirkung von H-NS mit ein, so lässt sich die erhöhte analoge Kontrolle in beiden Fällen dem Wildtyp zuordnen. (c) Verläufe der Differenzen der analogen CTCs für Schwellenparameter t zwischen 1 und 10000 auf logarithmischer Skala. Über den gesamten Bereich sinnvoller t -Werte (für $t > 10000$ geht die Zahl isolierter Knoten gegen 0, und $R \rightarrow \infty$) heben sich die *fis*-Datensätze (blau) von den *hns*-Datensätzen (gelb) ab.

4.3.3 Verteilung von Kontrolle

In den beiden vorangehenden Kapiteln wurden die Signaturen in den Expressionsprofilen mit der Verteilung der regulativen Kontrolle auf zwei verschiedene molekulargenetische Mechanismen erklärt. Das TRN vermittelt demnach digitale Regulation zwischen spezifischen Genen durch das Binden von Transkriptionsfaktoren. Im GPN werden Gene dagegen analog reguliert, benachbarte Loci reagieren konzertiert auf Änderungen der DNA-Struktur. Die Strukturproteine FIS und H-NS stabilisieren verdrehte DNA-Supercoids auf unterschiedlich Art und Weise. Während FIS das Binden der Polymerase und damit Transkription fördert, also aktivierend auf die entsprechenden Gene wirkt, unterdrückt das Binden von H-NS die Expression von Genen (Marr et al., 2007a). Abbildung 4.15 zeigt den Einfluss der beiden Komponenten auf den Wildtyp und die beiden Mutantenstämme, ermittelt aus den Inter- und Intrastammdaten der sieben Experimente.

Eine Interpretation der Expressionsdaten mit verschiedenen Kontrollmechanismen der Regulation hilft nicht nur bei der Analyse von Expressionsmustern, sie ist auch ein Beitrag zu einer Debatte in der Evolutionsbiologie. Dort wird seit einigen Jahren wieder verstärkt über Mechanismen diskutiert, die die Vererbung phenotypischer Eigenschaften erlauben. In komplexen Lebewesen ist eine solche Rückwirkung auf den Genotyp durch die Trennung von somatischen Zellen und Gameten ausgeschlossen. In Bakterien, wo diese Weismansche Barriere nicht existiert, ist ein Einfluss der Umgebung auf Veränderungen des Genotyps möglich. Nach Bateson (1963) würde ein solcher Einfluss jedoch die Flexibilität des Organismus gegenüber Umweltveränderung enorm einschränken. Die Spezialisierung von, beispielsweise, Bakterien würde demnach auf Mutationen führen, die in einer speziellen Umgebung zwar günstig, in jeder anderen aber ungünstig für die Zelle wären. Im Experiment jedoch zeigen *E. coli* Kulturen eine erstaunlich Flexibilität in ihrer genetischen Anpassungsfähigkeit (Novak et al., 2006). Mit der Verteilung von Kontrolle lässt sich die Kompensation von Mutationen auf die Regulation der Transkription erklären. Eine Spezialisierung an die Umgebung würde bei einer rein durch TFs vermittelten Regulationsmaschinerie zu einem Netzwerk führen, das weniger flexibel auf Änderungen reagieren kann. Die erhöhte digitale Kontrolle in den *fis*- und *hns*-Mutantenstämmen der Intrastamm-Experimente legt diesen Schluss nahe. Mittels der zweiten Kontrollkomponente, der analogen Regulation, ist der Organismus in der Lage, die Einschränkungen einer Mutation auszugleichen. Diese Fähigkeit zeigt sich in der erhöhten analogen Kontrolle des Wildtyps in den Interstamm-Experimenten.

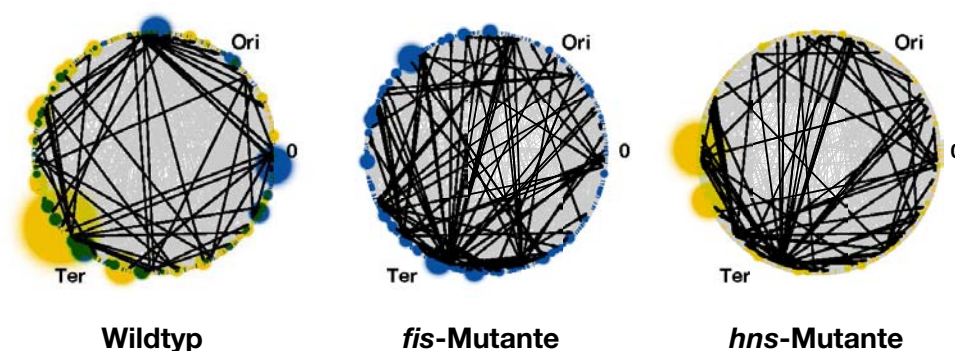


Abb. 4.15. Einfluss der analogen und digitalen Kontrolle auf die Expressionsmuster in Wildtyp, *fis*-Mutante und *hns*-Mutante. Schwarze Pfeile symbolisieren Wechselwirkungen des effektiven TRNs, ermittelt aus den Intrastamm-Daten. Graue Pfeile geben den Möglichkeitsraum der digitalen Regulation, das gesamte RegulonDB TRN an. Blau und gelbe Abschnitte auf dem zirkulären Genom zeigen differentiell exprimierte Gene in den Interstamm-Experimenten I_1 und I_2 (blau) und I_3 und I_4 (gelb). Die Zuweisung zu Wildtyp und Mutante ist im Text beschrieben. Die farbigen Kreisscheiben zeigen effektive GPN Komponenten an, der Radius der Scheiben skaliert mit dem Durchmesser der entsprechenden Subnetzwerke.

Ausblick

Die Beschreibung zellulärer Subsysteme mit Hilfe des Netzwerks der interagierenden Einzelemente hat sich in den letzten Jahren als eine wichtige Forschungsrichtung in der Analyse komplexer Systeme etabliert. Topologische Untersuchungen der Wechselwirkungsarchitektur haben das Verständnis bestimmter Systemfunktionen entscheidend mitgeprägt (siehe z.B. Shen-Orr et al., 2002; Wunderlich und Mirny, 2006; Carmi et al., 2006, zu Genregulation, Metabolismus, und Proteinwechselwirkungen). Ein Problem bei der Verallgemeinerung solcher Ansätze ist die begrenzte Datengrundlage biologischer Netze vieler Spezies. Die zunehmende Miniaturisierung der experimentellen Techniken und die Systematisierung biologischer Daten wird diesen Mangel in den nächsten Jahren aber voraussichtlich beheben. Eine generellere Herausforderung ist die Zusammenführung der Befunde verschiedener Disziplinen. Erst die Überwindung dieser künstlichen Aufteilung des Gesamtsystems – hervorgerufen durch die historisch gewachsene Abgrenzung einzelner Forschungsfelder – kann zu einem systemweiten Verständnis der Zelle und der Wechselwirkung von Zellverbänden in einem Organismus führen.

Wir wollen die Erforschung grundlegender Abhängigkeiten zwischen Topologie und Dynamik vorantreiben. Wie in Kapitel 1 gezeigt, enthält das Spektrum eines Graphen wertvolle Informationen über das topologische Potential, dass einem Prozess zur Verfügung steht. Trotz fundamentaler Einsichten in Abhängigkeiten zwischen topologischen Eigenschaften und den Spektren assoziierter Matrizen (z.B. Chung et al., 2003) ist der allgemeine Zusammenhang zwischen dem Charakter dynamischer Prozesse und der adäquaten Matrixrepräsentationen noch sehr unklar. Hypothesen, wie die einer Verbindung von konservativen Dynamiken mit der Laplace-Matrix und nicht-konservativen Dynamiken mit der Adjazenzmatrix, müssen anhand der Aussagekraft der Spektren überprüft werden. Auch das Verständnis der im Spektrum enthaltenen Information gilt es zu erweitern. Durch den Vergleich mit Ergebnissen der *random matrix theory* (Guhr et al., 1998), wo die Analyse von Spektren seit Jahrzehnten erfolgreich zur Interpretation physikalischer Systemeigenschaften herangezogen wird, hoffen wir, universelle Eigenschaften des topologischen Potentials von Netzwerken zu entdecken. Die Interpretation der Verteilung der Elemente im Perron-Frobenius-Eigenvektor (in Abschnitt 1.4) als Lokalisierungs/Delokalisierungs-Phänomen ist ein erster Schritt in diese Richtung.

Die Charakterisierung komplexer Netzwerke mit dynamischen Sonden ist ein neuartiger und vielversprechender Ansatz. Eine dynamische Sicht auf metabolische Prozesse ist unverzichtbar um Adaption und Flussreorganisation bei wechselnden Umweltbedingungen oder Mutationen zu verstehen. Das Ziel einer solchen Betrachtung ist die Kontrolle metabolischer Abläufe mit einer Vielzahl von Anwendungsmöglichkeiten. Dabei muss die Relevanz dynamischer Sonden mit den Ergebnissen der etablierten Modellvorstellungen verglichen werden. Die Grundvoraussetzungen solcher Analysen sind die Verwendung einer einheitlichen Topologie, die Wahl der richtigen Sonde, sowie der Vergleich geeigneter Observablen. Jüngste Arbeiten zu Fluktuationen in metabolischen Netzen (Müller-Linow et al., 2007) legen die Einbeziehung von Rauschen nahe – der Metabolismus könnte diese Fluktuationen nutzen, um besonders effizient auf Änderungen der externen Substratkonzentrationen zu reagieren. Der in Kapitel 2.4.1 entwickelte Formalismus bietet sich als Sonde hierfür an.

Ein attraktiver Gegenstand für diese Sonden könnte ein integriertes Genregulations-Netzwerk sein, das die Eigenschaften der spezifischen Bindungen von Transkriptionsfaktoren – des TRNs (siehe Abschnitt 4.1) – mit denen struktureller DNA-Veränderungen durch Supercoiling – des GPNs (siehe Abschnitt 4.3) – vereint. Ein solches Metanetzwerk, das die Eigenschaften verschiedener biologischer Regulationsmechanismen in einer Struktur verbindet, ähnelt topologisch den in Abschnitt 1.4 besprochenen Small-world-Graphen: Die auf der DNA benachbarten, kollektiv regulierten Genverbände sind durch Shortcuts mit entfernten Regionen auf dem Genom korreliert. Die spektrale Analyse eines solchen integrierten Netzwerks und die Annäherung an sich verändernde DNA-Strukturen (Postow et al., 2004) mit geeigneten zellulären Automaten versprechen Einsichten in evolutionär geprägte Strukturen dieser Art.

Das Ausmessen verschiedener biologischer Netze bezüglich ihres dynamischen Potentials mit dynamischen Sonden ist ein naheliegender nächster Schritt. Hypothesen, wie die einer dynamischen Optimierung von metabolischen und regulatorischen, nicht aber von Protein-Protein-Interaktionsnetzwerken, können mit dem im Abschnitt 2.1.2 entwickelten Apparat systematisch überprüft werden. Darüberhinaus kann mit der Implementierung einer Vielzahl von Regeln sehr unterschiedlichen Charakters, die Disposition einer Architektur zu einer bestimmten Dynamik offengelegt werden.

In Abschnitt 4.3 wurden die Genexpressionsmuster mit dem Zusammenspiel zweier Netzwerke, dem Transkriptions-Regulation-Netzwerk und dem Nachbarschaftsnetzwerk der Gene, interpretiert. Die Erklärungskraft dieses Ansatzes für genomweite Expressionsprofile muss, ähnlich wie in Abschnitt 4.2 für das TRN allein, evaluiert werden. Auf diese Weise lässt sich die Kontrollverteilung der Regulation in verschiedenen Expressionsmustern und damit in verschiedenen Zellzuständen bestimmen. Eine Zuweisung der Signaturen zu verstandenen und noch unbekannten Regulationsphänomenen ist dann möglich.

Mit der Analyse von Expressionsmustern geht die Modellierung der Regulationsmechanismen einher. Trotz existierender Modelle, wie den seit Jahrzehnten diskutierten Booleschen Netzwerken, und umfassenden Daten zu Regulationsnetzen einzelner Spezies, ist es noch nicht gelungen, die tatsächlichen Expressionsmuster von Microarray-Daten befriedigend zu simulieren. Eine Erweiterung der Booleschen Netze in mehrfacher Hinsicht ist dafür erforderlich. Der in Abschnitt 2.4.2 eingeführte Apparat, mit dem verrauschte Dynamiken auf beliebige Netzwerkstrukturen implementiert werden können, ist ein erster Schritt in diese Richtung. Darüberhinaus muss eine geeignete Übersetzung des binären Zustandsraums in Expressionsprofile entwickelt werden. Entscheidende Bedeutung kommt in diesem Zusammenhang dem Studium des Attraktorraums zu. Nur ein dynamisches System mit einer überschaubaren Zahl von Attraktoren, das Rauschen robust verrechnet – oder sogar nutzt –, und gleichzeitig sensibel auf relevante Signale reagiert, ist als Modell realer Prozesse geeignet. Die Kombination von Struktur und Dynamik ist für solche Ansätze von zentraler Bedeutung. Die Modellierung der Zelle als dynamisches System mit einem spezifischen Attraktorraum könnte helfen, wichtige Zellprozesse zu verstehen. Die Entschlüsselung der Regulationsmechanismen der Genexpression von Bakterien ist ein entscheidender Schritt auf dem Weg zu einer funktionalen Kontrolle des Zellprogramms.

Das Verständnis der prokaryotischen Regulationsmechanismen (wie aus Kapitel 4 resultierend) legt den Schritt zur Analyse eukaryotischer Expressionsmuster nahe. Ein möglicher Ansatz ist, sich der gewebe-spezifischen Expressionsdaten zu bedienen, um diese mit einem Modell der Genregulation abzugleichen und in einer selbstkonsistenten Methode mit den schon existierenden, disparaten Subnetzwerken der Genregulation des Menschen zusammenzufügen. Methoden, die genetische Netzwerke algorithmisch aus Expressionsdaten gewinnen, existieren bereits (Gardner et al., 2003). Ein so erzeugtes Regulationsnetzwerk des Menschen könnte das Verständnis komplexer genetischer Erkrankungen, bei denen das gleichzeitige Ausfallen mehrerer Gene dramatische Folgen für den Organismus hat, vertiefen.

Die Attraktivität der Netzwerktheorie ist eng verknüpft mit ihrer universellen Anwendbarkeit auf eine Vielzahl von Problemstellungen, Disziplinen und Größenordnungen. Auf einer molekulargenetischen Skala kann die Analyse von Regulationsnetzwerken helfen, neue epidemiologische Herausforderungen, wie beispielsweise Antibiotika-resistente Pestbakterien (Welch et al., 2007), anzunehmen. Auf einer viel größeren Skala kann eine Anwendung dynamischer Netzwerkmodelle helfen, Ausbreitungssignaturen einer Epidemie zu verstehen, um wirkungsvolle Konzepte der Prävention zu entwickeln. Die Bearbeitung solcher

Fragestellungen und die Beschreibung der evolutionären Ursprünge, Funktion und Dynamik komplexer biologischer Strukturen ist eine dringliche Aufgabe der theoretischen Biologie auf dem Weg hin zur Systembiologie.

Literaturverzeichnis

- Albert, R., 2005. Scale-free networks in cell biology. *J. Cell. Sci.* 118 (21), 4947–4957.
- Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47.
- Aldana, M., Coppersmith, S., Kadanoff, L. P., 2003. A celebratory volume in honor of Lawrence Sirovich. Springer Applied Mathematical Sciences Series, Ch. Boolean Dynamics with Random Couplings.
- Almaas, E., Oltvai, Z. N., Barabási, A.-L., 2005. The activity reaction core and plasticity of metabolic networks. *PLoS Comput. Biol.* 1 (7), e68.
- Amaral, L. A. N., Díaz-Guilera, A., Moreira, A. A., Goldberger, A. L., Lipsitz, L. A., 2004. Emergence of complex dynamics in a simple model of signaling networks. *Proc. Natl. Acad. Sci. USA* 101, 15551.
- Andrecut, M., Ali, M. K., 2001. Chaos in a simple Boolean network. *Int. J. Mod. Phys. B* 15 (1), 17–23.
- Arenas, A., Diaz-Guilera, A., Perez-Vicente, C. J., 2006. Synchronization reveals topological scales in complex networks. *Phys. Rev. Lett.* 96 (11), 114102.
- Arita, M., 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. USA* 101 (6), 1543–1547.
- Bak, P., Chen, K., Tang, C., 1990. A forest-fire model and some thoughts on turbulence. *Phys. Lett. A* 147 (5-6), 297–300.
- Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286, 509–512.
- Barrat, A., Weigt, M., 2000. On the properties of small-world network models. *Eur. Phys. J. B* 13, 547.
- Bateson, G., 1963. The role of somatic change in evolution. *Evolution* 17, 529–539.
- Blot, N., Mavathur, R., Geertz, M., Travers, A., Muskhelishvili, G., 2006. Homeostatic regulation of super-coiling sensitivity coordinates transcription of the bacterial genome. *EMBO reports* 7 (7), 710–715.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: Structure and dynamics. *Physics Reports* 424, 175–308.
- Bornholdt, S., 2005. Less is more in modeling large genetic networks. *Science* 310, 449–451.
- Carmi, S., Levanon, E. Y., Havlin, S., Eisenberg, E., 2006. Connectivity and expression in protein networks: Proteins in a complex are uniformly expressed. *Phys. Rev. E* 73, 031909.
- Cayley, A., 1857. On the theory of the analytical forms called trees. *Philos. Mag.* 13, 19–30.
- Cayley, A., 1875. On the Analytical Figures Called Trees in Mathematics and on Their Application in the Theory of Organic Compounds. *Berichte der deutschen chemischen Gesellschaft* 8, 1056–1059.
- Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., Herskowitz, I., 1998. The transcriptional program of sporulation in budding yeast. *Science* 282, 699–705.
- Chung, F., Lu, L., Vu, V., 2003. Spectra of random graphs with given expected degrees. *Proc. Natl. Acad. Sci. USA* 100 (11), 6313–6318.
- Costanzo, M. C., Crawford, M. E., Hirschman, J. E., Kranz, J. E., Olsen, P., Robertson, L. S., Skrzypek, M. S., Braun, B. R., Hopkins, K. L., Kondu, P., Lengieza, C., Lew-Smith, J. E., Tillberg, M., Garrels, J. I., 2001. YPD, PombePD and WormPD: model organism volumes of the BioKnowledge library, an integrated resource for protein information. *Nucleic Acids Res.* 29, 75–79.

- Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., Palsson, B. O., 2004. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429, 92–96.
- Csete, M., Doyle, J., 2004. Bow ties, metabolism and disease. *Trends Biotech.* 22 (9), 446–450.
- Davidson, E. H., Erwin, D. H., 2006. Gene Regulatory Networks and the Evolution of Animal Body Plans. *Science* 311 (5762), 796–800.
- Deutsch, A., Dormann, S., 2005. *Cellular Automaton Modeling and Biological Pattern Formation*. Birkhäuser, Boston.
- Dorman, C. J., 2004. H-NS: a universal regulator for a dynamic genome. *Nature Rev. Microbiol.* 2, 391–400.
- Drossel, B., 2005. Number of attractors in random boolean networks. *Phys. Rev. E* 72 (1), 016110.
- Drossel, B., Mihaljev, T., Greil, F., 2005. Number and Length of Attractors in a Critical Kauffman Model with Connectivity One. *Phys. Rev. Lett.* 94 (8), 88701.
- Drossel, B., Schwabl, F., 1992. Self-organized critical forest-fire model. *Phys. Rev. Lett.* 69, 1629.
- Edwards, J. S., Palsson, B. O., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* 97, 5528–5533.
- Erdős, P., Rényi, A., 1959. On random graphs. *Publ. Math. (Debrecen)* 6, 290.
- Ermentrout, G. B., Edelstein-Keshet, L., 1993. Cellular automata approaches to biological modeling. *J. Theor. Biol.* 160 (1), 97–133.
- Euler, L., 1736. Solutio problematis ad geometriam situs pertinentis. *Comment. Academiae Sci. I. Petropolitanae* 8, 128–140.
- Famili, I., Förster, J., Nielsen, J., Palsson, B. O., 2003. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl. Acad. Sci. USA* 100, 13134–13139.
- Farkas, I. J., Derényi, I., Barabási, A.-L., Vicsek, T., 2001. Spectra of 'real-world' graphs: Beyond the semicircle law. *Phys. Rev. E* 64, 026704.
- Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F., 1998. Stochastic resonance. *Rev. Mod. Phys.* 70, 223.
- Gardner, T. S., di Bernardo, D., Lorenz, D., Collins, J. J., 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301, 102–105.
- Giles, J., 2002. What kind of science is this? *Nature* 417, 216–218, news feature.
- Glasner, J. D., Liss, P., Plunkett III, G., Darling, A., Prasad, T., Rusch, M., Byrnes, A., Gilson, M., Biehl, B., Blattner, F. R., Perna, N. T., 2003. ASAP, a systematic annotation package for community analysis of genomes. *Nucleic Acids Res.* 31 (1), 147–151.
- Goh, K.-I., Kahng, B., Kim, D., 2001. Spectra and eigenvectors of scale-free networks. *Phys. Rev. E* 64, 051903.
- Graham, I., Matthai, C. C., 2003. Investigation of the forest-fire model on a small-world network. *Phys. Rev. E* 68, 036109.
- Gray, L., 2003. A Mathematician Looks at Wolfram's New Kind of Science. *Notices of AMS* 50 (2), 200–211.
- Guelzim, N., Bottani, S., Bourguin, P., Képès, F., 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31, 60.
- Guhr, T., Müller-Groeling, A., Weidenmüller, H. A., 1998. Random-matrix theories in quantum physics: common concepts. *Physics Reports* 299, 189.
- Gutierrez-Rios, R. M., Rosenblueth, D. A., Loza, J. A., Huerta, A. M., Glasner, J. D., Blattner, F. R., Collado-Vides, J., 2003. Regulatory Network of *Escherichia coli*: Consistency Between Literature Knowledge and Microarray Profiles. *Genome Res.* 13 (11), 2435–2443.
- Harary, F., 1969. *Graph Theory*. Addison-Wesley, Reading, Massachusetts.
- Hayes, B., 2002. The World According to Wolfram. *American Scientist* 90, 308–312.
- Herrgard, M. J., Covert, M. W., Palsson, B. O., 2003. Reconciling Gene Expression Data With Known Genome-Scale Regulatory Network Structures. *Genome Res.* 13 (11), 2423–2434.
- Hubbel, C. H., 1965. An input-output approach to clique identification. *Sociometry* 28, 377–399.
- Hufnagel, L., Brockmann, D., Geisel, T., 2004. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. USA* 101 (42), 15124–15129.

- Huss, M., Holme, P., 2006. Currency and commodity metabolites: Their identification and relation to the modularity of metabolic networks, e-print q-bio.MN/0603038.
- Hütt, M.-T., Marr, C., 2006. Selbstorganisation als metatheorie. In: Selbstorganisation. Ein Denksystem für Natur und Gesellschaft. Böhlau Verlag, Köln Weimar Wien, pp. 106–126.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., Barabási, A.-L., 2000. The large-scale organization of metabolic networks. *Nature* 407, 651–654.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28 (1), 27–30.
- Kauffman, K. J., Prakash, P., Edwards, J. S., 2003a. Advances in flux balance analysis. *Curr. Opin. Biotech.* 14, 491–496.
- Kauffman, S., Peterson, C., Samuelsson, B., Troein, C., 2003b. Random boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. USA* 100, 14796–14799.
- Kauffman, S., Peterson, C., Samuelsson, B., Troein, C., 2004. Genetic networks with canalizing boolean rules are always stable. *Proc. Natl. Acad. Sci. USA* 101, 17102–17107.
- Kauffman, S. A., 1969. Metabolic stability and epigenesis in randomly constructed nets. *J. Theor. Biol.* 22, 437.
- Kirchhoff, G., 1847. über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer ströme geführt wird. *Ann. Phys. Chem.* 72, 497–508.
- Kitano, H., 2002. Systems Biology: A Brief Overview. *Science* 295, 1662–1664.
- Kitano, H., 2004. Biological robustness. *Nature Rev. Genet.* 5, 862.
- Kleinberg, J. M., 1999. Authoritative Sources in a Hyperlinked Environment. *J. ACM* 46, 604–632.
- Klemm, K., Bornholdt, S., 2005. Stable and unstable attractors in Boolean networks. *Phys. Rev. E* 72 (5), 55101.
- Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H., 2005. *Systems Biology in Practice*. Wiley, Weinheim.
- Kramer, B., Ohtsuki, T., Kettemann, S., 2005. Random Network Models and Quantum Phase Transitions in Two Dimensions. *Physics reports* 417 (5-6), 211–342.
- Kuramoto, Y., 1984. *Chemical Oscillations, Waves and Turbulence*. Springer, Berlin.
- Langton, C. G., 1990. Computation at the edge of chaos: phase transitions and emergent computation. *Physica D* 42, 12.
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., B. Ren, Wyrick, J. J., J.-B. Tagne, Volkert, T. L., Fraenkel, E., Gifford, D. K., Young, R. A., 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799.
- Li, F., Long, T., Lu, Y., Ouyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. USA* 101, 4781–4786.
- Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M., Teichmann, S. A., Gerstein, M., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 431, 308–312.
- Lynch, A. S., Lin, E. C. C., 1996. Responses to molecular oxygen. In: *Escherichia coli and Salmonella: cellular and molecular biology*, 2nd Edition. American Society for Microbiology, Washington DC, pp. 1526–1538.
- Ma, H., Zeng, A.-P., 2003a. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423–1430.
- Ma, H., Zeng, A.-P., 2003b. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19, 270–277.
- Ma, H.-W., Buer, J., Zeng, A.-P., 2004a. Hierarchical structure and modules in the *escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics* 5, 199.
- Ma, H. W., Zhao, X. M., Yuan, Y. J., Zeng, A. P., 2004b. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics* 20 (12), 1870–1876.

- Marguerat, S., Jensen, T. S., de Lichtenberg, U., Wilhelm, B. T., Jensen, L. J., Bähler, J., 2006. The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast. *Yeast* 23 (4), 261–277.
- Marr, C., Geertz, M., Hütt, M.-T., Muskhelishvili, G., 2007a. Mutations of chromatin proteins reveal distinct logical types of transcriptional control, eingereicht.
- Marr, C., Hütt, M.-T., 2005. Topology regulates pattern formation capacity of binary cellular automata on graphs. *Physica A* 354, 641–662.
- Marr, C., Hütt, M.-T., 2006. Similar impact of topological and dynamic noise on complex patterns. *Phys. Lett. A* 349, 302–305.
- Marr, C., Müller-Linow, M., Hütt, M.-T., 2007b. Regularizing capacity of metabolic networks. *Phys. Rev. E* 75, im Druck.
- Marr, C., Shehadeh, L. A., Liebovitch, L. S., Hütt, M.-T., 2007c. Influence of global statistical patterns of gene regulatory networks on microarray data, eingereicht.
- Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. *Science* 296, 910–913.
- Matache, M. T., Heidel, J., 2004. Random boolean network model exhibiting deterministic chaos. *Phys. Rev. E* 69 (5), 056214.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple building blocks of complex networks. *Science* 298, 824–827.
- Moreira, A. A., Mathur, A., Diermeier, D., Amaral, L. A. N., 2004. Efficient system-wide coordination in noisy environments. *Proc. Natl. Acad. Sci. USA* 101, 12085.
- Müller-Linow, M., Marr, C., Hütt, M.-T., 2006. Topology regulates the distribution pattern of excitations in excitable dynamics on graphs. *Phys. Rev. E* 74, 016112.
- Müller-Linow, M., Weckwerth, W., Hütt, M.-T., 2007. Consistency of metabolic correlation networks with genomederived pathway structures, unpublished.
- Newman, M. E. J., 2000. Models of the Small World. *J. Stat. Phys.* 101, 819–841.
- Newman, M. E. J., 2003a. Mixing patterns in networks. *Phys. Rev. E* 67, 026126.
- Newman, M. E. J., 2003b. The structure and function of complex networks. *SIAM Rev.* 45, 167.
- Newman, M. E. J., 2006. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74 (3), 036104.
- Novak, M., Pfeiffer, T., Lenski, R. E., Sauer, U., Bonhoeffer, S., 2006. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. *Am. Nat.* 168, 242–251.
- Oltvai, Z. N., Barabási, A.-L., 2002. Life's Complexity Pyramid. *Science* 298, 763–764.
- Page, L., Brin, S., Motwani, R., Winograd, T., 1998. The pagerank citation ranking: Bringing order to the web. Stanford University Technical Report.
- Palsson, B., 2006. Systems biology: properties of reconstructed networks. Cambridge University Press, New York.
- Pastor-Satorras, R., Vespignani, A., 2001. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 86, 3200.
- Peak, D., West, J. D., Messinger, S. M., Mott, K. A., 2004. Evidence for complex, collective dynamics and emergent, distributed computation in plants. *Proc. Natl. Acad. Sci. USA* 101, 918.
- Pecora, L. M., Carroll, T. L., 1998. Master stability functions for synchronized coupled systems. *Phys. Rev. Lett.* 80 (10), 2109–2112.
- Postow, L., Hardy, Christine D. and Arsuaga, J., Cozzarelli, N. R., 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* 18 (14), 1766–1779.
- Restrepo, J. G., Ott, E., Hunt, B. R., 2006. Characterizing the dynamical importance of network nodes and links. *Phys. Rev. Lett.* 97 (9), 094102.
- Salgado, H., Gama-Castro, S., Peralta-Gil, M., Diaz-Peredo, E., Sanchez-Solano, F., Santos-Zavaleta, A., Martinez-Flores, I., Jimenez-Jacinto, V., Bonavides-Martinez, C., Segura-Salazar, J., Martinez-Antonio, A., Collado-Vides, J., 2006. RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.* 34, D394–397.

- Samuelsson, B., Troein, C., 2003. Superpolynomial Growth in the Number of Attractors in Kauffman Networks. *Phys. Rev. Lett.* 90 (9), 98701.
- Sawai, S., Thomason, P. A., Cox, E. C., 2005. An autoregulatory circuit for long-range self-organization in *Dictyostelium* cell populations. *Nature* 433, 323.
- Segré, D., Vitkup, D., Church, G. M., 2002. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 99, 15112–15117.
- Shannon, C. E., 1948. A mathematical theory of communication. *The Bell Systems Technical Journal* 27, 379–423.
- Shen-Orr, S. S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31, 64–68.
- Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. USA* 102 (21), 7695–7700.
- Socolar, J. E. S., Kauffman, S. A., 2003. Scaling in Ordered and Critical Random Boolean Networks. *Phys. Rev. Lett.* 90 (6), 68702.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E. D., 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190–193.
- Strogatz, S. H., 2001. Exploring complex networks. *Nature* 410, 268–276.
- Tanaka, R., 2005. Scale-rich metabolic networks. *Phys. Rev. Lett.* 94 (16), 168101.
- Travers, A., Muskhelishvili, G., 2005a. Bacterial chromatin. *Curr. Opin. Genet. Dev.* 15, 507–514.
- Travers, A., Muskhelishvili, G., 2005b. DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.* 3, 157–169.
- Trusina, A., Maslov, S., Minnhagen, P., Sneppen, K., 2004. Hierarchy measures in complex networks. *Phys. Rev. Lett.* 92, 178702.
- Tu, B. P., Kudlicki, A., Rowicka, M., McKnight, S. L., 2005. Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science* 310, 1152–1158.
- Turing, A., 1952. The Chemical Basis of Morphogenesis. *Phil. Trans. R. Soc. London B* 237, 37–72.
- von Neumann, J., 2001. The general and logical theory of automata. In: Taub, A. H. (Ed.), *J. von Neumann, Collected Works. Vol. 5*. Macmillan, New York, p. 288.
- Wagner, A., Fell, D. A., 2001. The small world inside large metabolic networks. *Proc. R. Soc. B* 268, 1803.
- Watts, D. J., Strogatz, S. H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.
- Welch, T. J., Fricke, W. F., McDermott, P. F., White, D. G., Rosso, M., et al., 2007. Multiple antimicrobial resistance in plague: An emerging public health risk. *PLoS One* 2(3), e309.
- Wigner, E. P., 1967. Random matrices in physics. *SIAM Rev.* 9 (1), 1–23.
- Wolfram, S., 1983. Statistical mechanics of cellular automata. *Rev. Mod. Phys.* 55, 601.
- Wolfram, S., 1984. Universality and complexity in cellular automata. *Physica D* 10, 1.
- Wolfram, S., 2002. A new kind of science. Wolfram Media, Champaign, IL.
- Wunderlich, Z., Mirny, L. A., 2006. Using the Topology of Metabolic Networks to Predict Viability of Mutant Strains. *Biophys. J.* 91 (6), 2304.
- Yoshimoto, H., Saltsman, K., Gasch, A. P., Li, H. X., Ogawa, N., Botstein, D., Brown, P. O., Cyert, M. S., 2002. Genome-wide Analysis of Gene Expression Regulated by the Calcineurin/Crz1p Signaling Pathway in *Saccharomyces cerevisiae*. *J. Bio. Chem.* 277 (34), 31079–31088.
- Yu, H., Gerstein, M., 2006. Colloquium Papers: Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. USA* 103, 14724–14731.

Zusammenfassung

Netzwerke sind für das Verständnis vieler biologischer Prozesse von elementarer Bedeutung. Die Anwendung graphentheoretischer Methoden erstreckt sich auf Systeme sehr verschiedener Größenordnungen: von der Beschreibung ineinandergreifender Genregulationsmechanismen über die Architektur metabolischer Netze bis hin zu der Ausbreitung von Infektionskrankheiten.

Diese Arbeit untersucht den Zusammenhang zwischen Topologie und Dynamik auf zwei Ebenen. Zum einen werden dynamische Sonden entwickelt und zur Analyse von Netzwerktopologien eingesetzt. Eine dieser Sonden vereinheitlicht zwei gut untersuchte, dynamische Modelle – zelluläre Automaten und Boolesche Netze – für beliebige Systemarchitekturen. Auf abstrakten, synthetischen Graphen lassen sich so die dynamischen Konsequenzen topologischer Modifikationen nachvollziehen. Für reale metabolische Netzwerke zeigt sich, dass die globale Struktur der Stoffwechselreaktionen die Regularisierung komplexer Dynamik befördert – unabhängig von der spezifischen Netzwerkrepräsentation und der verwendeten Dynamik. Stationäre Metabolitenkonzentrationen, die sich in fast jedem Zellmodus in kurzer Zeit einregeln, werden so aus einer dynamischen Sicht auf das Gesamtsystem erklärt.

Zum anderen verwenden wir Konzepte und Werkzeuge aus der Graphentheorie, um die Signaturen dynamischer Prozesse auf dem Genom von *E. coli* zu interpretieren. Das zentrale Ergebnis dieser Untersuchung ist die verteilte Kontrolle transkriptioneller Genregulation: Unter bestimmten Bedingungen wird sie verstärkt durch spezifische Transkriptionsfaktoren (digital), unter anderen Bedingungen durch die Veränderung der räumlichen DNA-Struktur (analog) vermittelt. Diese Eigenschaft des Regulationsapparates weist auf eine besondere genetische Anpassungsfähigkeit der Bakterien gegenüber Umweltveränderungen hin. Einschränkungen der Flexibilität der digitalen Komponente durch Adaption und, folglich, Mutationen können durch den regulierenden Einfluss der DNA-Strukturproteine, der analogen Komponente, ausgeglichen werden.

Lebenslauf

Geburtsdatum	28. Januar 1977
Geburtsort	Würzburg
Nationalität	deutsch
Familienstand	ledig, zwei Kinder

Ausbildung

1983 - 1987	Grundschule Burghaslach
1987 - 1996	Gymnasium Scheinfeld, mathematisch-naturwissenschaftlicher Zweig, Abschluss: Abitur mit 1,2
1997 - 2002	Technischen Universität München, Studium der Allgemeinen Physik, Spezialfächer Quantenoptik und Halbleiterphysik, Abschluss: Diplom mit 1,3
2001 - 2002	Max-Planck-Institut für Quantenoptik, Garching, Diplomarbeit <i>Zustandspräparation in offenen Quantensystemen</i> in der Quantendynamik Gruppe von Prof. Dr. G. Rempe
2004 - 2007	Graduiertenstipendium der Technischen Universität Darmstadt, Doktorarbeit <i>Dynamische Prozesse auf abstrakten Graphen und biologischen Netzwerken</i> in der Theoretische Biologie und Bioinformatik Gruppe von Prof. Dr. M. Hütt

Wissenschaftliche Praxis

April 2001	Forschungsaufenthalt am Institut Laue-Langevin, Grenoble, Frankreich. Klassifizierung neuer Neutronendetektoren
Januar - März 2003	Forschungsaufenthalt am Imperial College, London, in der Quantuminformations Gruppe von Prof. Dr. P. L. Knight, Design von Quantengattern
2003 - 2004	Berlin-Brandenburgischen Akademie der Wissenschaften, Wissenschaftliche Hilfskraft in der Arbeitsgruppe Selbstorganisation der Jungen Akademie. Wissenschaftliche Hilfskraft an der TU Darmstadt
Juli - August 2005	Forschungsaufenthalt am Center for Complex Systems and Brain Sciences, Florida Atlantic University, Florida, USA. Analyse von Gen-chip Expressionsmustern und Transkriptions-Regulations Netzwerken in der Gruppe von Prof. Dr. L. Liebovitch

Publikationen

- C. Marr, A. Beige & G. Rempe, 2003. Entangled-state preparation via dissipation-assisted adiabatic passages. *Phys. Rev. A* 68 (3), 33817.
- A. Beige, H. Cable, C. Marr & P. L. Knight, 2005. Speeding up gate operations through dissipation. *Laser physics* 15 (1), 162–169.
- C. Marr & M.-T. Hütt, 2005. Topology regulates pattern formation capacity of binary cellular automata on graphs. *Physica A* 354, 641–662.
- C. Marr & M.-T. Hütt, 2006. Similar impact of topological and dynamic noise on complex patterns. *Phys. Lett. A* 349, 302–305.
- M. Müller-Linow, C. Marr & M.-T. Hütt, 2006. Topology regulates the distribution pattern of excitations in excitable dynamics on graphs. *Phys. Rev. E* 74, 016112.
- M.-T. Hütt & C. Marr, 2006. Selbstorganisation als metatheorie. In: *Selbstorganisation. Ein Denksystem für Natur und Gesellschaft*. Böhlau Verlag, Köln Weimar Wien, S. 106–126.
- C. Marr, M. Müller-Linow & M.-T. Hütt, 2007. Regularizing capacity of metabolic networks. *Phys. Rev. E* 75, im Druck.
- C. Marr, M. Geertz, M.-T. Hütt & G. Muskhelishvili, 2007. Mutations of chromatin proteins reveal distinct logical types of transcriptional control, eingereicht.
- C. Marr, L. A. Shehadeh, L. S. Liebovitch & M.-T. Hütt, 2007. Influence of Global Statistical Patterns of Gene Regulatory Networks on Microarray Data, eingereicht.
- M. Zakhartsev, D. Petroi, P. Dimitrov, R. Khusainov, C. Marr, M. Hütt & H. O. Pörtner, 2007. Metabolic rearrangements compensate effect of temperature acclimation in yeast *Saccharomyces cerevisiae*, eingereicht.

Konferenzbeiträge

69. Jahrestagung der Deutschen Physikalischen Gesellschaft, Berlin, 4.-9. März 2005. Poster Präsentation: C. Marr & M.-T. Hütt. Topology regulates pattern formation capacity of binary cellular automata on graphs.
- Dynamics on Complex Networks and Applications. Max-Planck-Institut für Physik komplexer Systeme, Dresden, 12.-17. Februar 2006. Poster Präsentation: C. Marr, M. Müller-Linow, & M.-T. Hütt. Metabolic networks are enhanced for robust dynamic performance.
- EMBO Workshop on Principles of Self-Organization in Living Matter. EMBL Heidelberg, 2.-5. Juni 2006. Poster Präsentation: C. Marr & M.-T. Hütt. Noise processing in complex networks.
- Bremen Molecular and Marine Biology Meeting, Schloss Etelsen, 26.-27. Januar 2007. Poster Präsentation: N. Sonnenschein, C. Marr, & M.-T. Hütt. Analyzing effects of in silico gene deletions on flux distributions and information flow in metabolic networks.

Dank

an

Marc Hütt

für das Vertrauen und die stetige Motivation,

Georgi Muskhelishvili, Larry Liebovitch, Lina Shehadeh, Marcel Geertz, Mark Müller-Linow
für die inspirierende Zusammenarbeit,

Manuel Dehnert, Niko Sonnenschein, Daniel Geberth, Jens Raderschall, Martin Heindel und
Anne Christmann

für wertvolle Kommentare.

Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

Darmstadt, den 16. April 2007